Migration Center User Guide

Best Practices

Issue	01
Date	2025-03-14

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process.* For details about this process, visit the following web page:
https://www.huawei.com/en/psirt/vul-response-process
For vulnerability information, enterprise customers can visit the following web page:
https://securitybulletin.huawei.com/enterprise/en/security-advisory

Contents

1 Configuring Permissions Required for Server Migration

Overview

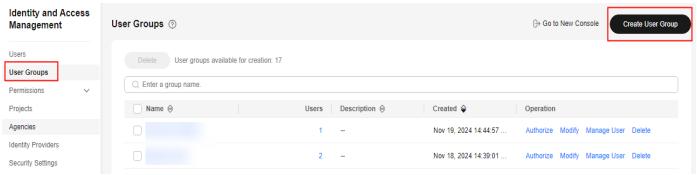
1. Create a user group named **migration_users** and assign the permissions required to use MgC and SMS to the user group. The IAM user to be created will inherit the permissions from the user group.
2. For a user in the local **admin** group, create an IAM user who is named **mgc-user**, belongs to the **migration_users** user group, and has only programmatic access to Huawei Cloud. The IAM user is not allowed to access the Huawei Cloud console using a password.
3. Provide the MgC Agent with the AK/SK pair generated when **mgc-user** is created. The AK/SK pair is used to register the MgC Agent with MgC and authenticate API calling during the migration.

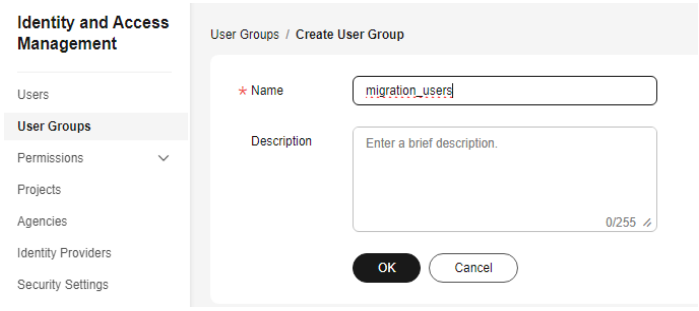
Step 1: Create a User Group

Step 1 Log in to the **IAM console**.

Step 2 On the IAM console, choose **User Groups** from the left navigation pane, and click **Create User Group** in the upper right corner.

Figure 1-1 Creating a user group



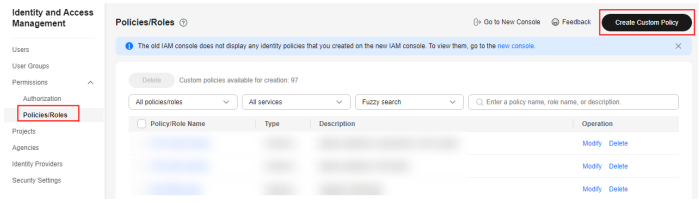


----End

Step 2: Create a Permissions Policy

Step 1 On the IAM console, in the navigation pane, choose **Permissions** > **Policies/Roles** and click **Create Custom Policy** in the upper right corner.

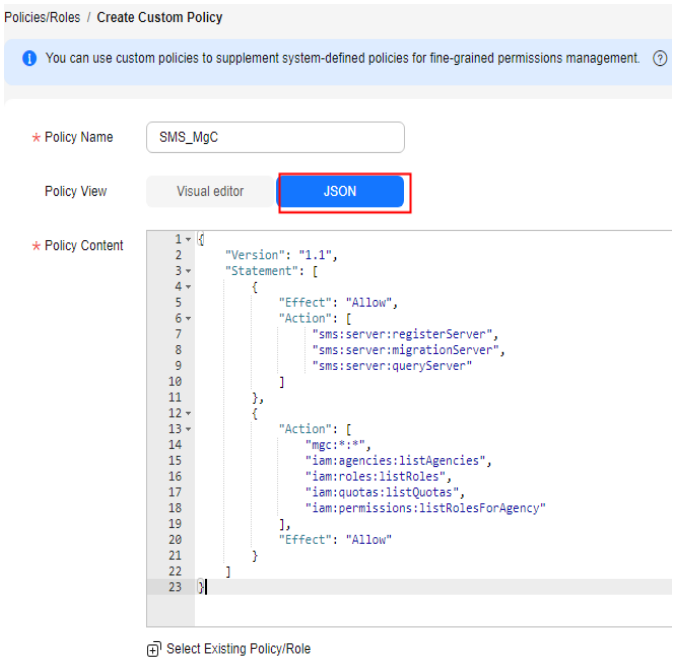
Figure 1-2 Creating a custom policy



Step 2 Create a policy for using SMS, a global cloud service. Enter a policy name, set **Policy View** to **JSON**, and copy the following content to the **Policy Content** box.

```
{
   "Version": "1.1",
   "Statement": [
      {
         "Effect": "Allow",
         "Action": [
               "sms:server:registerServer",
               "sms:server:migrationServer",
               "sms:server:queryServer"
         ]
      },
      {
         "Action": [
            "mgc:*:*",
            "iam:agencies:listAgencies",
            "iam:roles:listRoles",
            "iam:quotas:listQuotas",
            "iam:permissions:listRolesForAgency"
         ],
         "Effect": "Allow"
      }
   ]
}
```


Figure 1-3 Creating a policy that defines the permissions required for using SMS



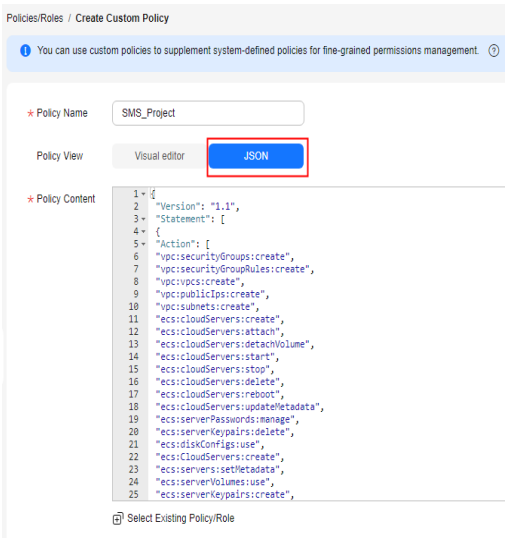
Step 3 Create a policy for using regional cloud services that SMS depends on. Enter a policy name, set **Policy View** to **JSON**, and copy the following content to the **Policy Content** box.

```
{
 "Version": "1.1",
 "Statement": [
 {
 "Action": [
 "vpc:securityGroups:create",
 "vpc:securityGroupRules:create",
 "vpc:vpcs:create",
 "vpc:publicips:create",
 "vpc:subnets:create",
 "ecs:cloudServers:create",
 "ecs:cloudServers:attach",
 "ecs:cloudServers:detachVolume",
 "ecs:cloudServers:start",
 "ecs:cloudServers:stop",
 "ecs:cloudServers:delete",
 "ecs:cloudServers:reboot",
 "ecs:cloudServers:updateMetadata",
 "ecs:serverPasswords:manage",
 "ecs:serverKeypairs:delete",
 "ecs:diskConfigs:use",
 "ecs:CloudServers:create",
 "ecs:servers:setMetadata",
 "ecs:serverVolumes:use",
 "ecs:serverKeypairs:create",
 "ecs:serverInterfaces:use",
 "ecs:serverGroups:manage",
 "ecs:securityGroups:use",
 "ecs:servers:unlock",
 "ecs:servers:rebuild",
 "ecs:servers:lock",
 "ecs:servers:reboot",
 "evs:volumes:use",
 "evs:volumes:create",
 "evs:volumes:update",
 "evs:volumes:delete",
 "evs:volumes:attach",
```



```
"evs:volumes:detach",
"evs:snapshots:create",
"evs:snapshots:delete",
"evs:snapshots:rollback",
"ecs:*:get*",
"ecs:*:list*",
"evs:*:get*",
"evs:*:list*",
"vpc:*:list*",
"vpc:*:get*",
"ims:*:get*",
"ims:*:list*"
],
"Effect": "Allow"
}
]
}
```

Figure 1-4 Creating a policy for using the regional cloud services that SMS depends on



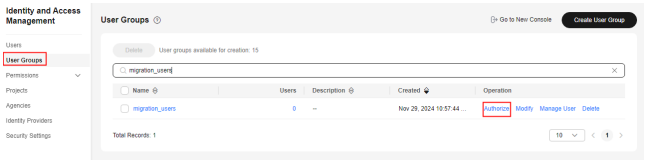
----End

Step 3: Assign Permissions

Step 1 On the IAM console, choose **User Groups** from the navigation pane.

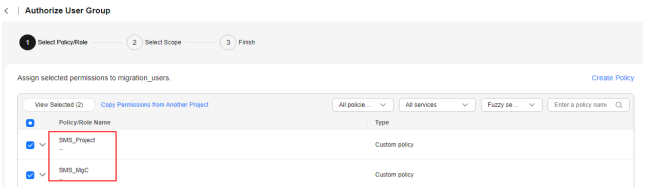
Step 2 In the user group list, locate the user group created in **step 1** and click **Authorize** in the **Operation** column.

Figure 1-5 Assigning permissions to the user group



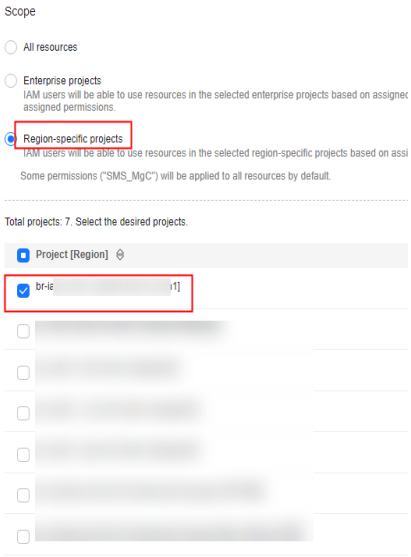
Step 3 Search for and select the two custom policies created in **step 2** and click **Next**.

Figure 1-6 Selecting the created custom policies



Step 4 Select **Region-specific projects** for **Scope** and select a region-specific project. Then the IAM users in the group can use resources in the region-specific project based on their permissions.

Figure 1-7 Selecting a region-specific project



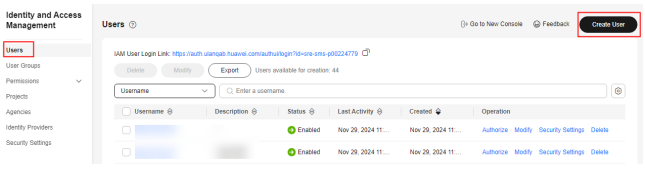
Step 5 Click **OK**.

----End

Step 4: Create a User

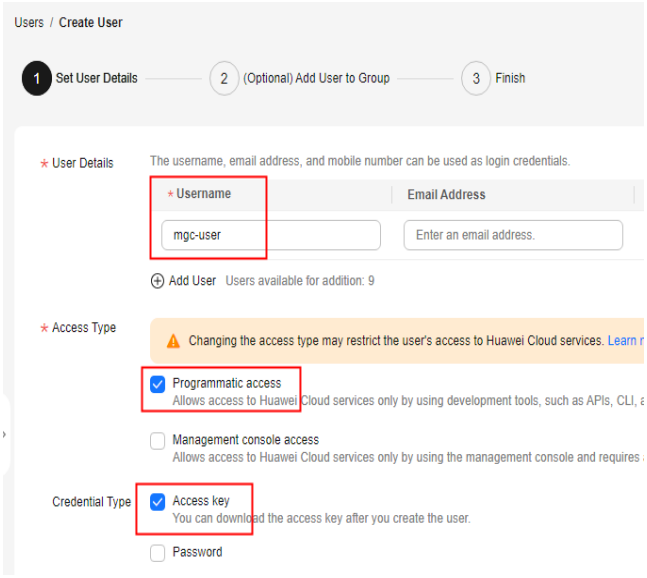
Step 1 On the IAM console, choose **Users** from the left navigation pane, and click **Create User** in the upper right corner.

Figure 1-8 Creating a user



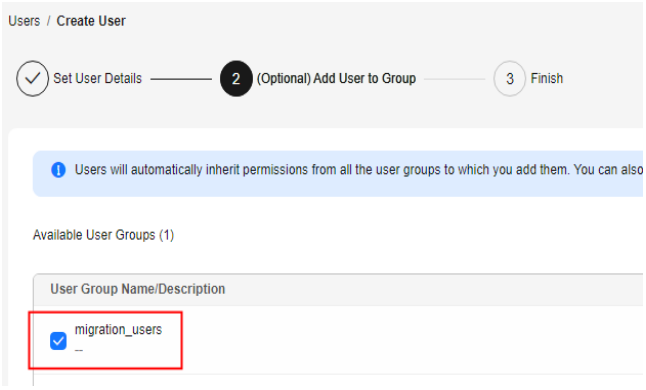
Step 2 Enter a username, deselect **Management console access**, and click **Next**.

Figure 1-9 Configuring basic information



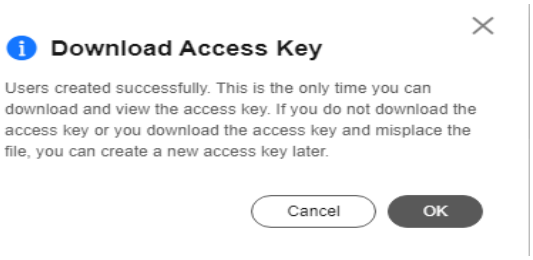
Step 3 Select the user group created in **step 1** and click **Create**.

Figure 1-10 Selecting a user group



Step 4 After the user is created, the **Download Access Key** dialog box is displayed. Click **OK** to download an AK/SK pair for the IAM user.

Figure 1-11 Downloading an access key



----End

2 Server Migration

2.1 Network Requirements for Server Migration

Background

A server migration involves two types of traffic: control flow and data flow.

- **Control flow** refers to the communication between the source server and cloud service management planes. To ensure a smooth migration, verify if there are any restrictions on the outbound traffic from the source server. Additionally, confirm that the source server can access the following cloud services via their domain names: SMS, IAM, ECS, EVS, VPC, and IMS.
- **Data flow** refers to data transmission from the source server to the target server. To ensure smooth data transmission, confirm that the security group of the target server allows traffic from the source server's IP address over the specified migration ports. If the source server cannot directly access the Internet or cannot communicate with the target server, a proxy server must be configured. In this case, consider the following factors:
  - The proxy server can correctly forward traffic from the source server.
  - The proxy server's security group is configured to allow traffic from the source server's IP address over the proxy port.

For General Migration Scenarios

Internet access is required for migration using MgC.

1. Install the MgC Agent in the intranet environment. The MgC Agent must be able to access the source servers to be migrated. For details, see **Installing the MgC Agent on Windows**.
2. Complete steps 2 to 5 described in **Migrating On-premises Servers to Huawei Cloud Using MgC**.

> **CAUTION**
>
> Ensure that the security group of the target server allows access from the source servers over the specified migration ports. For details about how to configure security group rules, see **How Do I Configure Security Group Rules for Target Servers?**

3. **Create a server migration workflow**.
   - If the source servers can access the Internet, set the migration network to **Public**.

     Figure 2-1 Migration over the Internet

   - If the source servers cannot access the Internet, prepare a proxy server that can access the Internet and install proxy software. For details, see step 1 in **How Do I Configure a Source Server to Access Huawei Cloud Through a Proxy?** The network requirements for the proxy server are as follows:
     - Regardless whether the proxy server is on the source intranet environment or on the cloud, it must be able to access the source servers to be migrated over an intranet.
     - The proxy server must be configured to allow inbound traffic from the source servers over the proxy port.

     After the proxy server is configured, set the migration network to **Private**, and enter the private IP address of the proxy server and the port specified for the proxy software.

     Figure 2-2 Migration over a private network

2.2 Migrating On-premises Servers to Huawei Cloud

Scenario

This section describes how to use MgC to migrate on-premises servers to Huawei Cloud.

Preparations

- Prepare a Windows server for installing the MgC Agent in the source intranet environment. The Windows server must:

  - Be able to access the Internet and the domain names of MgC, IoTDA, and other cloud services. For details about the domain names to be accessed, see **Domain Names**.
  - Use PowerShell **3.0** or later.
  - Have at least 4 CPUs and 8 GB of memory.
  - Allow outbound traffic on port 8883 if the server is in a security group.
  - Not have any antivirus or protection software enabled. This type of software may stop the MgC Agent from executing migration commands, resulting in migration failures.

⚠ CAUTION

Do not install the MgC Agent on a source server to be migrated.

- **High resource consumption**: The MgC Agent consumes CPU and memory resources during collection and migration. If a large number of migration tasks are performed by the MgC Agent, services on the source server may be affected.
- **Port occupation**: The MgC Agent occupies some ports on the server, which may affect services running on it.

- The Windows server where the MgC Agent is installed must be able to access the source servers you want to migrate over the following ports:
  - Windows: port 5985
  - Linux: port 22
- WinRM must be enabled on Windows source servers, and these source servers must be able to access the server where the MgC Agent is installed. For more information, see **How Do I Configure WinRM and Troubleshoot WinRM Connection Problems?**
- Migrating Windows source servers requires a Windows agent image. Since Huawei Cloud no longer provides public Windows images due to security restrictions, you need to prepare a Windows agent image on your own before the migration. For details, see **Using Custom Agent Images**.
- Prepare a Huawei account or an IAM user that can access MgC. For details, see **Preparations**.
- **Create a migration project** on the MgC console.

Notes

Before creating a server migration workflow, read and understand the following precautions.

Item	Precaution
Source download bandwidth	Used to download SMS-Agent to source servers. • If each source server uses a dedicated bandwidth, the bandwidth must be at least 30 Mbit/s. • If source servers share a bandwidth, the average bandwidth must be at least 50 Mbit/s.
Migration bandwidth	• Used to migrate data. It affects the migration speed and duration. • For details about how to estimate the migration duration, see **How Long Does a Migration Take?**
CPU and memory	• At least 520 MB of available memory • At least 0.3 CPUs available in Linux and at least 1 CPU available in Windows
OS compatibility	For details about what OSs are supported, see **Supported OSs**.
Server migration statements	For details about the important statements you need to understand before the migration, see **What Are the Important Statements of SMS?**
Notes and constraints	For details about the notes and constraints for server migration, see **Notes and Constraints**.
Billing	For details about the fees that may be incurred during the migration, see **Billing**.
Permissions configuration	For details about the permissions the target account must have, see **Permissions Management**.
Migration network and ports	For details about the requirements for the migration network and ports, see **How Do I Set Up a Secure Migration Network for Using SMS?**

Step 1: Download and Install the MgC Agent

Install the MgC Agent and connect it to MgC. For more information, see **Installing the MgC Agent**.

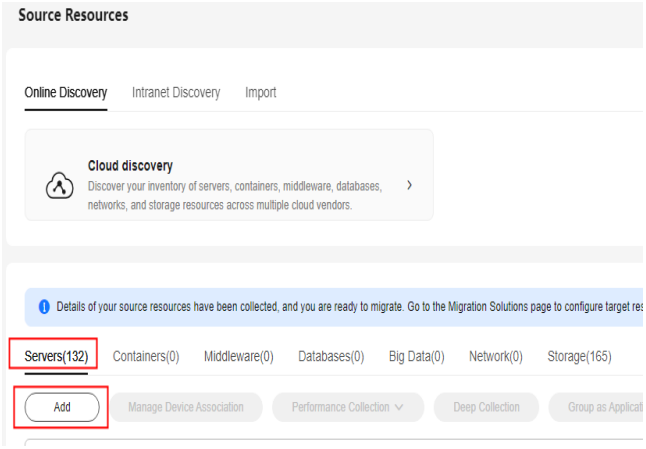
Step 2: Add Servers to MgC

Step 1 Sign in to the **MgC console**. In the navigation pane, under **Project**, select your **application migration project** from the drop-down list.

Step 2 In the navigation pane, choose **Discover** > **Source Resources**.

Step 3 On the **Servers** tab, click **Add** above the list.

Figure 2-3 Add Servers to the Application



Step 4 In the displayed dialog box, configure parameters listed in **Table 2-1** and click **Confirm**. The system automatically checks the credential status and starts collecting resource details.

Table 2-1 Parameters for adding a server

Parameter	Description
Name	User-defined
MgC Agent	Select the MgC Agent installed in the source environment.
Type	Select the OS type of the source server.
Access IP Address	Enter the IP address of the source server. If the source server is in the same VPC as the MgC Agent, you can enter the private IP address of the server. Otherwise, you have to enter its public IP address.
Port	Enter the port on the source server that allows access from the MgC Agent. • By default, port 5985 on Windows source servers must be opened to the MgC Agent. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the MgC Agent. You can specify a different port if needed.
Credential	Select the server credential. If the credential is not displayed in the list, go to the MgC Agent console, **add the server credential**, and synchronize it to MgC.

Step 5 After the server is added, view it in the list.

----End

Step 3: Group Servers as an Application

You can group the added servers as an application to get sizing recommendations for target resources and execute the migration.

Step 1 On the **Resources** page, in the **Servers** list, select the servers to be grouped as an application and click **Group as Application** above the list.

Step 2 Select an application from the drop-down list. If no applications are available, click **Create Application**. In the displayed dialog box, enter an application name and description; select a business scenario, environment, and target region; and click **Create**. For more information, see **Creating an Application**.

Step 3 Click **OK**. You can view the application name in the **Application** column of these servers.

----End

(Optional) Step 4: Associate Source Servers with Existing Servers on Huawei Cloud

If you have servers on Huawei Cloud, you can associate source servers with these existing Huawei Cloud servers. These Huawei Cloud servers will be used to receive data migrated from their paired source servers. Then you can go to **Step 6: Create a Migration Workflow** directly.

If you do not want to migrate data to these existing Huawei Cloud servers, skip the current step and go to **Step 5: Create an Application Assessment**.

CAUTION

Before associating an existing server on Huawei Cloud with a source server, make sure that the existing server meets the following requirements:

- Disks on the existing server can be formatted. During the migration, disks on the existing server will be formatted and re-partitioned based on the source disk settings for receiving data migrated from the source server.
- To migrate over the Internet, the existing server must be able to access the Internet.
- The existing server must be in the same region as the **application** that the source server is added to.

Step 1 In the navigation pane on the left, choose **Migration Solutions**.

Step 2 Click **View Resources** in the **Target Configuration** card.

Step 3 On the displayed **Servers** tab, locate a source server and click **Associate** in the **Target Association** column.

Step 4 In the displayed dialog box, select the region of the **application** and select a project. Then, select an existing Huawei Cloud server and click **Confirm**.

After the association is complete, **Associated** appears in the **Target Association** column. You can click **Details** to view the specifications of the associated target server.

----End

Step 5: Create an Application Assessment

Assessing an application can get recommendations for most suitable Huawei Cloud resources based on the specifications, performance, and business purpose data of the source resources added to the application, as well as your selected recommendation references, such as, cost or performance reference and ECS type references.

NOTE

If your source servers have been **associated with existing servers** on Huawei Cloud, you can skip this step and create a migration workflow to migrate them.

Step 1 On the **Migration Solutions** page, click **Assess** in the **Target Configuration** card.

Step 2 In the **Select Application** drop-down list, select the **application** that contains the source servers to be assessed.

Step 3 In the **Select Resources** area, select the servers to be assessed.

Step 4 Configure an assessment policy based on **Table 2-2**.

Table 2-2 Settings used for computing target recommendations

Parameter	Option	Description
Target Region	-	Select the region where you want to purchase resources on Huawei Cloud. You are advised to select a region close to your target users for lower network latency and quick access.
Assessment Policy	Match source configuration	MgC will recommend Huawei Cloud resources in the same or slightly larger size as source resources. For details about how MgC recommends appropriate target resources, see **How Does MgC Generate Target Recommendations?**
	Match business scenario	MgC recommends the right Huawei Cloud resources based on the business scenario of source resources and Huawei Cloud best practices. For details about how MgC recommends appropriate target resources, see **How Does MgC Generate Target Recommendations?**
Priority	High performance	MgC recommends target resources with optimal performance.
	Low cost	MgC recommends the most cost-effective target resources that meet your demands.

Parameter	Option	Description
Preferences	Server Types (Optional)	Select the server types you prefer.
	Server Series (Optional)	Select the server series you prefer. The system will generate recommendations based on your preferred server types and series. **NOTICE** If you select **Display only series allowed on DeHs**, **Server Types** will be dimmed, and the server series allowed on DeHs in the target region will be listed.
	System Disk (Optional)	Select the system disk type you prefer.
	Data Disk (Optional)	Select the data disk type you prefer.
	Sizing Criteria	Choose the criteria that the system will use to generate server recommendations. For details about how MgC recommends appropriate target resources, see **How Does MgC Generate Target Recommendations?**

Step 5 Click **OK**.

Step 6 In the application list on the **Migration Solutions** page, locate the applications and click **View Target Configurations** in the **Operation** column.

In the **Target Configurations** area, you can view the specifications of Huawei Cloud resources recommended based on the source resource specifications and your preferences. It also gives you the ability to estimate the cost of running on Huawei Cloud. In addition, you can modify the recommended target configurations.

----End

Step 6: Create a Migration Workflow

After **step 1** to **step 5** are complete, create a workflow to migrate the source servers to Huawei Cloud. For details, see **Creating a Server Migration Workflow**.

2.3 Migrating Servers from Alibaba Cloud to Huawei Cloud

2.3.1 Overview

This best practice describes the detailed procedure and precautions for migrating servers from Alibaba Cloud to Huawei Cloud.

The key steps include:

1. **Making Preparations**: Ensure that the migration accounts are available and the accounts have required permissions.
2. **Installing the MgC Agent**: Download and install the MgC Agent, a tool provided by MgC.
3. **Discovering Resources**: Collect information about the Alibaba Cloud ECSs to be migrated.
4. **Assessing Migration Readiness**: Check the configuration of the Alibaba Cloud ECSs to be migrated, test their network connectivity, and gather them into groups.
5. **Assessing Target Servers**: Get recommendations for Huawei Cloud resources and configure target servers for source servers.
6. **Creating a Workflow**: Create a migration workflow to migrate your source servers.

Notes

Before creating a server migration workflow, read and understand the following precautions.

Item	Precaution
Source download bandwidth	Used to download SMS-Agent to source servers. • If each source server uses a dedicated bandwidth, the bandwidth must be at least 30 Mbit/s. • If source servers share a bandwidth, the average bandwidth must be at least 50 Mbit/s.
Migration bandwidth	• Used to migrate data. It affects the migration speed and duration. • For details about how to estimate the migration duration, see **How Long Does a Migration Take?**
CPU and memory	• At least 520 MB of available memory • At least 0.3 CPUs available in Linux and at least 1 CPU available in Windows
OS compatibility	For details about what OSs are supported, see **Supported OSs**.
Server migration statements	For details about the important statements you need to understand before the migration, see **What Are the Important Statements of SMS?**
Notes and constraints	For details about the notes and constraints for server migration, see **Notes and Constraints**.

Item	Precaution
Billing	For details about the fees that may be incurred during the migration, see **Billing**.
Permissions configuration	For details about the permissions the target account must have, see **Permissions Management**.
Migration network and ports	For details about the requirements for the migration network and ports, see **How Do I Set Up a Secure Migration Network for Using SMS?**

2.3.2 Preparations

To ensure a smooth migration, you need to complete the following preparations:

Preparing a Huawei Account

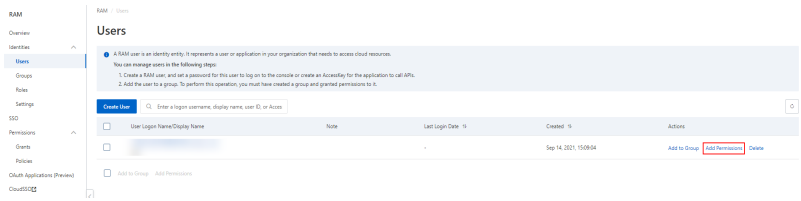
Before using MgC, prepare a HUAWEI ID or an IAM user that can access MgC and obtain an access key (AK/SK) of the account or IAM user. For more information, see **Preparations**.

Obtaining an Access Key for Your Alibaba Cloud

Check whether your Alibaba Cloud account has an AK/SK pair and has the **AliyunECSReadOnlyAccess** permissions. If it does not have, perform the following steps to generate the AK/SK pair and add the required permissions for it:

1. Sign in to the RAM console using your Alibaba Cloud account.
2. In the navigation pane on the left, choose **Identities** > **Users**.
3. On the **Users** tab, click **Create User**.
4. On the **Create User** page, in the **User Access Key** area, click **Create Access Key**.
5. Assign permissions to the RAM user.

   On the **Users** page, click **Add Permissions** in the **Operation** column and grant the **AliyunECSReadOnlyAccess** permissions to the RAM user.



Creating an Application Migration Project

Create a migration project on the MgC console. For details, see **Managing Migration Projects**.

2.3.3 Step 1: Download and Install the MgC Agent

The MgC Agent is a migration tool provided by MgC. It is used to discover source resources and execute migration commands from MgC.

Procedure

Step 1 Prepare a Windows server in the source intranet environment and install the MgC Agent on the server. For details about the requirements for the server and how to install the MgC Agent, see **Installing the MgC Agent on Windows**.

⚠ CAUTION

Do not install the MgC Agent on a source server to be migrated.

- **High resource consumption**: The MgC Agent consumes CPU and memory resources during collection and migration. If a large number of migration tasks are performed by the MgC Agent, services on the source server may be affected.
- **Port occupation**: The MgC Agent occupies some ports on the server, which may affect services running on it.

Step 2 Log in to the MgC Agent console and connect the MgC Agent to MgC. For details, see **Connecting the MgC Agent to MgC**.

Step 3 After the connection is successful, add the credentials of the source servers to be migrated to the MgC Agent. For details, see **Adding Resource Credentials**. Correctly configure **Resource Type** based on the source servers when you add their credentials.

----End

2.3.4 Step 2: Discover Alibaba Cloud ECSs

Prerequisites

- You have completed all **preparations**.
- You have **installed the MgC Agent** in the source environment and connected it to MgC.

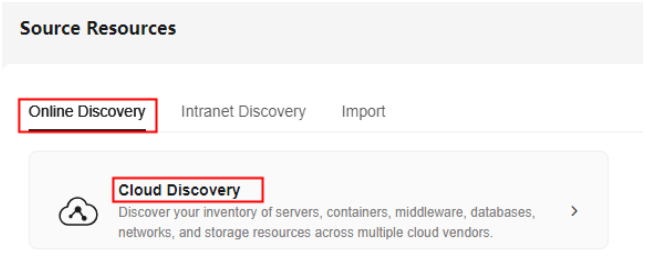
Procedure

Step 1 Sign in to the **MgC console**. In the navigation pane, under **Project**, select your **application migration project** from the drop-down list.

Step 2 In the navigation pane, choose **Source Resources**.

Step 3 Under **Online Discovery**, click **Cloud Discovery**.

Figure 2-4 Cloud discovery



Step 4 Configure the parameters based on **Table 2-3**.

Table 2-3 Parameters for creating an Internet-based discovery task

Region	Parameter	Description	Mandatory
Basic Settings	Task Name	Enter a task name.	Yes
	Task Description	Describe the task.	No
Task Settings	Source Platform	Select **Alibaba Cloud**.	Yes
	Credential	Select the credential for accessing Alibaba Cloud. If the credential has not been added, choose **Create** to add it. For details, see **Managing Credentials**. **NOTICE** Select **AK/SK** for **Authentication** and enter the AK/SK pair of your Alibaba Cloud account. Your account must have the **AliyunECSReadOnlyAccess** permissions.	Yes
	Region	Select the region where the source servers are located. You can select multiple regions.	Yes
Resource Discovery	Cloud Platform Collection	Enable cloud platform collection, select **Servers** from the **Resource Type** drop-down list.	Yes

Region	Parameter	Description	Mandatory
	Application Association (Optional)	An application is a group of resources that need to be migrated together. You can add resources to or remove resource from an application as needed. You can use the application to get recommendations for target resources and create a workflow to migrate the source resources. • If an application is available, select the application from the **Application** drop-down list. • If no application is available, click **Create Application**. In the displayed dialog box, enter an application name and description, select a business scenario, environment, and region (where you are migrating to), and click **OK**.	No

Step 5 Click **Confirm**. After the task is created, the system automatically starts collecting information about your Alibaba Cloud ECSs.

- On the **Source Resources** page, click **View** next to **Total tasks** to go to the task list. You can view the task status and task details. If the task status is **Failed**, click **View** in the **Operation** column to view the data source that failed to be collected. You can move the cursor to the collection status of the data source to view the failure cause.
- On the **Source Resources** page, in the server list, click **Server** in the **Category** column or the number in the **Total Resources** column. On the **Servers** tab, you can view the discovered source servers and their details.

----End

2.3.5 Step 3: Assess Migration Readiness

Measure whether the source servers are ready to migrate from the items of basic configuration, network environment, and migration group.

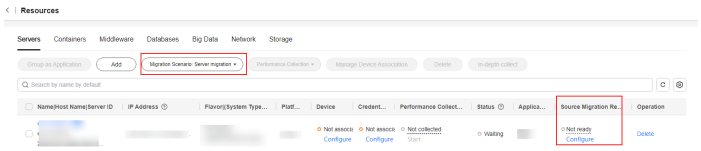
Prerequisites

- Your **Alibaba Cloud ECSs have been discovered**.
- You have **added source server credentials** to the MgC Agent.

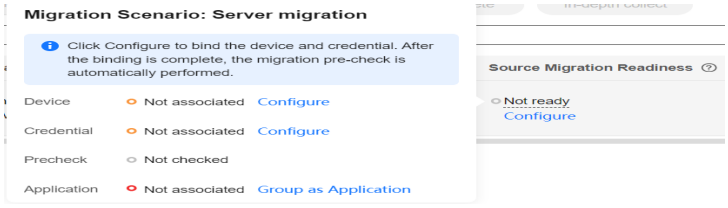
Procedure

Step 1 On the **Source Resources** page, click the **Servers** tab.

Step 2 On the top of the server list, choose **Migration Scenario** > **Server migration**.



Step 3 Locate a source server and move the cursor to **Not ready** in the **Migration Readiness** column. You can view the configurations that need to be completed to make the server ready. You need to associate the server with an MgC agent and a credential, pass the migration pre-check (automatically triggered), and add the server to an application. Click **Configure** in the **MgC Agent** or **Credential** column.



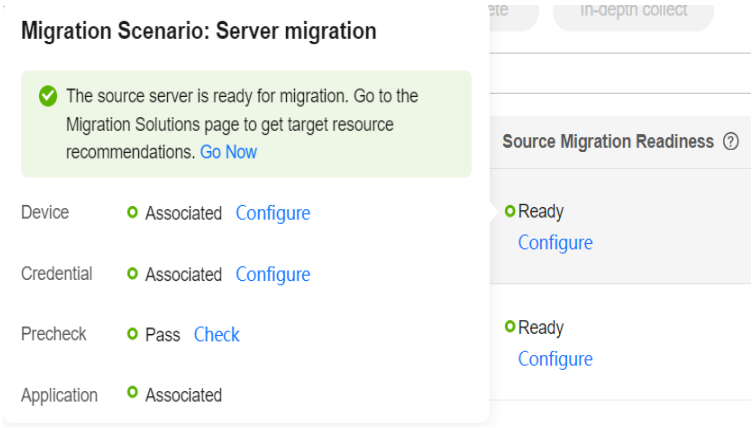
Step 4 Configure the parameters listed in **Table 2-4**.

Table 2-4 Parameters for configuring migration readiness

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
MgC Agent	Select the **MgC Agent installed in the source environment**. If there are a large number of servers to be associated with this the MgC Agent, you can select these servers and click **Manage MgC Agent** above the list to associate them in a batch.
IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration. During the migration, the MgC Agent accesses the source server through this IP address.
Port	Select the source server's port that allows traffic from the MgC Agent. • By default, port 5985 on Windows source servers must be opened to the MgC Agent. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the MgC Agent. You can specify a different port if needed.
Credential	Select the server credential you added to the MgC Agent. If the credential is not displayed in the list, go to the MgC Agent console, **add the server credential**, and synchronize it to MgC.

Step 5 Click **Confirm**. The system checks whether the source server can be accessed from the MgC Agent using the information you provided, and collect resource details again if necessary. The pre-check takes about 10 seconds.

Step 6 Add the source server to an application. For details, see **Grouping Resources as Applications**. Check whether **Ready** is displayed in the **Migration Readiness** column.



CAUTION

To reduce migration risks, you are advised to group no more than 30 servers as an application. If more than 30 servers need to be migrated, group them as multiple applications.

----End

2.3.6 Step 4: Create an Application Assessment

Assessing an application can generate recommendations for rightsized Huawei Cloud resources based on the specifications, performance, and business purpose of the source resources added to the application. These recommendations also take into account your requirements for cost, availability, and compliance.

Prerequisites

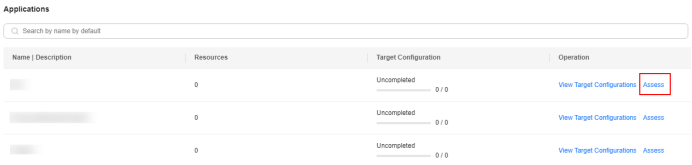
You have **assessed the migration readiness** of source resources and grouped them as applications.

Procedure

Step 1 Sign in to the **MgC console**. In the navigation pane, under **Project**, select your **application migration project** from the drop-down list.

Step 2 In the navigation pane, choose **Migration Solutions**. On the **Migration Solutions** page, you can view the list of applications created in the current project.

Step 3 In the application list, locate the application you want to assess and click **Assess** in the **Operation** column.



Step 4 In the **Select Resources** area, select the servers to be assessed in the application.

Step 5 Configure an assessment policy based on **Table 2-5**.

Table 2-5 Settings used for computing target recommendations

Parameter	Option	Description
Target Region	-	Select the region where you want to purchase resources on Huawei Cloud. You are advised to select a region close to your target users for lower network latency and quick access.
Assessment Policy	Match source configuration	MgC will recommend Huawei Cloud resources in the same or slightly larger size as source resources. For details about how MgC recommends appropriate target resources for you, see **How Does MgC Generate Target Recommendations?**
	Match business scenario	MgC recommends the right Huawei Cloud resources based on the business scenario of source resources and Huawei Cloud best practices. For details about how MgC recommends appropriate target resources for you, see **How Does MgC Generate Target Recommendations?**
Priority	High performance	MgC recommends target resources with optimal performance.
	Low cost	MgC recommends the most cost-effective target servers that meet your demands.
Preferences	Server Types (Optional)	Select the server types you prefer.
	Server Series (Optional)	Select the server series you prefer. The system will generate recommendations based on your preferred server types and series. **NOTICE** If you select **Display only series allowed on DeHs**, **Server Types** will be dimmed, and the server series allowed on DeHs in the target region will be listed.
	System Disk (Optional)	Select the system disk type you prefer.
	Data Disk (Optional)	Select the data disk type you prefer.

Parameter	Option	Description
	Sizing Criteria	Choose the criteria that the system will use to generate server recommendations. For details about how MgC recommends appropriate target resources for you, see **How Does MgC Generate Target Recommendations?**

Step 6 Click **Create Assessment**. After the assessment task is complete, you can **view the assessment results** which include the recommended specifications of target resources. You can also **view server performance data**.

Step 7 (Optional) Perform the following operations:

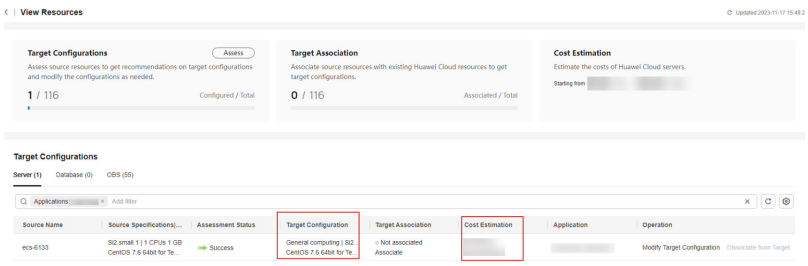
- **Modify target recommendations**. You can modify the recommended specifications for target servers and their disks.
- **Associate source servers with target servers**. If you already have servers that match your requirements on Huawei Cloud, you can associate them with source servers.

----End

Viewing Target Recommendations

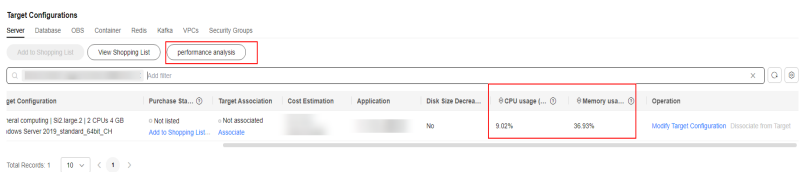
In the application list on the **Migration Solutions** page, click **View Target Configurations** in the **Operation** column.

In the **Target Configurations** area, you can view the specifications of Huawei Cloud resources recommended based on the source resource specifications and your preferences. It also gives you the ability to estimate what it will cost to run your services on Huawei Cloud.



Viewing Server Performance Data

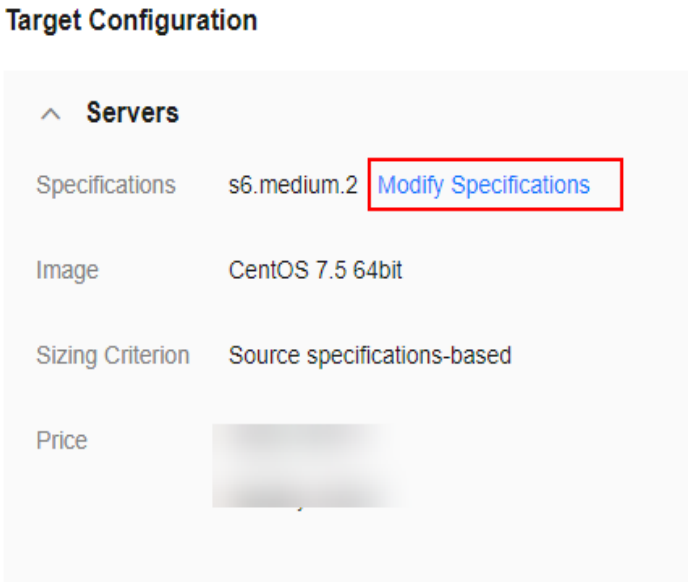
On the **Target Configurations** page, in the server list, you can view the average CPU and memory usage of each server over the last 7 or 30 days. Click **Performance Analysis** to view the performance statistics of all servers.



Modifying Target Recommendations

Step 1 In the **Target Configurations** area, locate the server that you want to modify the recommended target configurations for and click **Modify Target Configuration** in the **Operation** column.

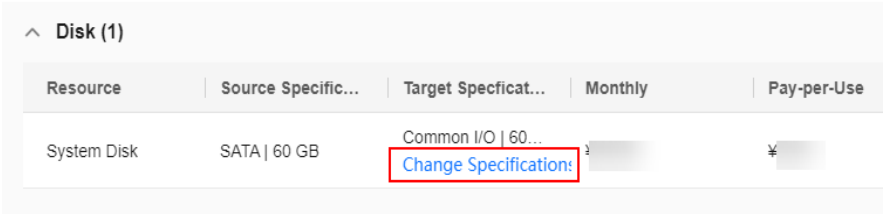
Step 2 Modify the specifications and image for the target server.



Step 3 In the disk area, locate a disk and click **Modify Specifications** in the **Target Specifications** column. You can modify the disk type and capacity. Only disks on Linux target servers can be downsized if the paired source servers have over-provisioned storage resources. If you downsize a disk for the target server, the system will set **Disk Downsized** to **Yes**. The reverse also applies.

NOTICE

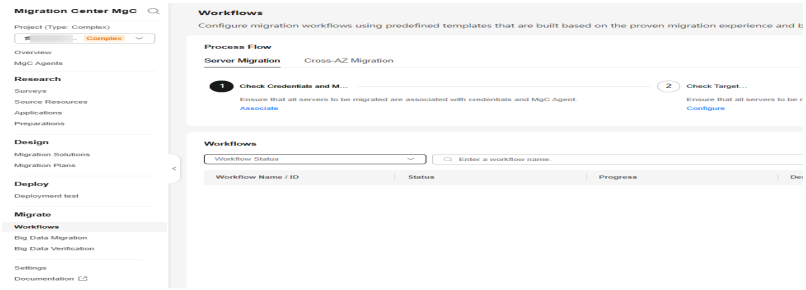
- The system disk capacity ranges from 40 GB to 1,024 GB.
- The data disk capacity ranges from 10 GB to 32,768 GB.
- Disk downsizing is only available for Linux, and the decreased sizes must be larger than the used sizes of the source disks.
- In the cross-AZ migration scenario, only disk upsizing is supported. Even if you choose to downsize disks here, the settings will not be applied, and the system will create target disks as large as source disks.



----End

2.3.7 Step 5: Create a Server Migration Workflow

After all operations in **Preparations** and **Step 1** to **Step 4** are complete, you can go to the **Workflows** page to **create a server migration workflow**.



2.4 One-stop Cross-AZ ECS Migration

Scenario

Use MgC to quickly migrate cloud servers from one AZ to another. This practice applies to migration of fewer than 30 ECSs in a single batch across AZs within a region. You only need to specify a resource group name, and MgC takes care of all the rest, from resource discovery, collection, and assessment to migration.

Preparations

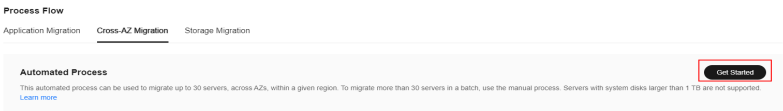
You need to prepare a Huawei account or an IAM user that can access MgC. For details, see **Preparations**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Overview**.

Step 3 In the **Process Flow** area, click the **Cross-AZ Migration** tab. In the **Automated Process** area, click **Get Started**.



Step 4 In the displayed dialog box, specify an application name and select the target AZ you want to migrate to.

Step 5 Click **Create and Run**. MgC will automatically collect information about servers in the selected source AZ under the current account, creates an application, adds the discovered servers to the application, and starts the assessment process.

Step 6 After the assessment process is complete, click **Close** to configure the workflow.

Step 7 Configure the workflow parameters listed in **Table 2-6**.

Table 2-6 Parameters required for configuring a workflow

Area	Parameter	Description
Workflow Details	Name	Enter a workflow name.
	Description	Enter a workflow name.
Application	Application	Select the application defined in **Step 4**.
Migration Settings	Region	Select the region where the source AZ is located. The region configured in the application is populated by default.
	Target AZ	Select the AZ you want to migrate to. The configuration must be the same as that of the created application.
	Target Network	Only **Retain original** is available.
	Target Server	Create now. MgC creates backups and images for source servers, and uses the images to create target servers immediately after the workflow runs.
	Stop Target Server	• If you select **Yes**, target servers will be stopped after being created. • If you select **No**, target servers will be started after being created.
	Stop Source Server	• If you select **Yes**, source servers will be stopped before incremental backups are created for them. This ensures data consistency as high as possible. • If you select **No**, source servers remain running when incremental backups are created for them.
	Create System Disk Image	• If you select **Yes**, a system disk image will be created for each of the source servers. The images can be used to reinstall the OS for the paired target servers. • If you select **No**, the system will not create system disk images for the source servers.
Advanced Settings	Delete Intermediate Resources	If this function is enabled, intermediate resources generated during the migration, such as backups, snapshots, and images, will be deleted after the service cutover is complete.

Area	Parameter	Description
	Retain Primary NIC IP Addresses	If this function is enabled, the private and public IP addresses of the primary NIC on source servers will be retained on target servers, and random private IP addresses will be allocated to source servers. If a rollback is needed, it has to be performed manually.

Step 8 Configure the workflow and click **Next: Confirm**. After confirming that the configuration is correct, click **Create**. The migration workflow will be created and displayed in the workflow list.

CAUTION

After a migration workflow is created, it switches to the **Waiting** status, and the migration has not started.

Step 9 Click the workflow name to go to the details page. The steps are predefined steps in the template. You can **add stages and steps** to the workflow.

Step 10 To start the migration, click **Run** in the **Operation** column.

- You can view the migration progress on the **Steps** tab. The workflow can continue only after you perform the manual steps contained.
- On the **Servers** tab, you can view the migration status of each server.

----End

2.5 Migrating Servers Across AZs on Huawei Cloud

Scenario

This section describes how to use MgC to migrate a large number of servers between AZs within a region of Huawei Cloud. For a small-scale, single-batch migration of fewer than 30 servers, see **One-stop Cross-AZ ECS Migration**.

Preparations

- Prepare a Huawei account or an IAM user that can access MgC. For details, see **Preparations**.
- **Create a migration project** on the MgC console.

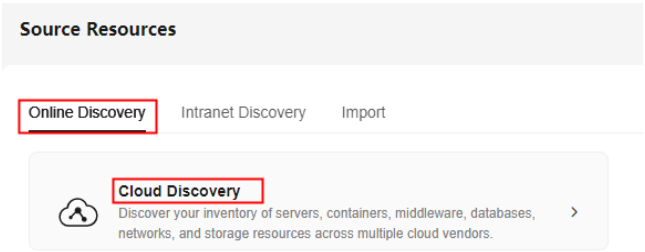
Step 1: Discovers Servers in the Source AZ

Step 1 Sign in to the **MgC console**. In the navigation pane, under **Project**, select your **application migration project** from the drop-down list.

Step 2 In the navigation pane, choose **Source Resources**.

Step 3 Under **Online Discovery**, click **Cloud Discovery**.

Figure 2-5 Cloud discovery



Step 4 Configure the parameters listed in **Table 2-7**.

Table 2-7 Parameters for creating an Internet-based discovery task

Region	Parameter	Description	Mandatory
Task Basics	Task Name	Enter a task name.	Yes
	Task Description	Describe the task.	No
Task Settings	Source Platform	Select **Huawei Cloud**.	Yes
	Credential	Select the credential of the source account. If no credential is available, choose **Create** to create a credential by referring to **Adding a Credential**. **NOTE** The AK/SK pair of the source account must be specified in the new credential.	Yes
	Region	Select the region where the source servers are located. You can select multiple regions.	Yes

Step 5 Enable cloud platform collection, select **Servers** from the **Resource Type** drop-down list.

Step 6 (Optional) Group the servers to be discovered as an application.

- If an application is available, select the **application** from the **Application** drop-down list.
- If no application is available, click **Create Application**. In the displayed dialog box, enter an application name and description, select **Cross-AZ migration** for **Business Scenario**, select the target region and AZ, and click **OK**.

Step 7 Click **OK**. After the discovery task is created, MgC starts to automatically discover servers in the selected regions selected in **Step 4**.

- On the **Source Resources** page, in the resource list, click **Server** in the **Category** column or the number in the **Total Resources** column.
- On the **Source Resources** page, in the **Discovery Tasks** card, click **View** next to **Total tasks**. If the task status is **Failed**, click **View** in the **Operation** column to view the data source that failed to be collected. You can move the cursor to the collection status of the data source to view the failure cause.

----End

Step 2: Group Servers as an Application

If the servers discovered have already been grouped into an application in **step 6**, ski p this section and go to **Step 3: Getting Target Recommendations**.

Step 1 In the **Servers** list of the **Resources** page, select the servers to be added to the same application and choose **Resource Management** > **Manage Application Association** in the upper left corner.

Step 2 Select the application from the drop-down list. If no application is available, click **Create Application**. In the displayed dialog box, enter an application name and description, select **Cross-AZ migration** for **Business Scenario**, select the target region and AZ, and click **OK**. For details, see **Creating an Application**.

Step 3 Click **OK**. You can view the application name in the **Application** column of these servers.

----End

Step 3: Getting Target Recommendations

Step 1 On the **Migration Solutions** page, click **Assess** in the **Target Configuration** card.

Step 2 In the **Select Application** drop-down list, select the **application** that contains the source servers to be assessed.

Step 3 In the **Select Resources** area, select the resources to be assessed in the application.

Step 4 Configure an assessment policy based on **Table 2-8**.

Table 2-8 Settings used for computing target recommendations

Parameter	Description
Target Region	Select the region you want to migrate to.
Assessment Policy	Select **Cross-AZ migration** and select the target AZ.
Priority	• High performance MgC recommends target resources with optimal performance. • Low cost MgC recommends the most cost-effective target resources that meet your demands.

Parameter	Description
Preferences	You can select server types, server series, and disk types you prefer. Your preferences have the highest priority during the resource assessment.

Step 5 Click **OK**.

Step 6 In the application list on the **Migration Solutions** page, locate the applications and click **View Target Configurations** in the **Operation** column.

In the **Target Configurations** area, you can view the specifications of Huawei Cloud resources recommended based on the source resource specifications and your preferences. It also gives you the ability to estimate the cost of running on Huawei Cloud. In addition, you can modify the recommended target configurations.

----End

Step 4: Creating a Cross-AZ Migration Workflow

After **step 1** to **step 3** are complete, **create a cross-AZ migration workflow**.

2.6 Keeping Private IP Addresses of Servers Unchanged After the Migration

In MgC server migration workflows, you can choose to retain private IP addresses for source servers on target servers after the migration. This feature can reduce the need to modify service code due to IP address changes.

Disclaimer

- **Service availability risks**

  This feature can ensure that the private IP addresses of source servers are retained on target servers. It does not guarantee your services can run properly on the target servers. You need to evaluate and assume the risks arising from using this feature.

- **Rollback description**

  Migration workflows cannot automatically roll back the IP addresses of target servers to their original ones. If any problems happen when you use this function, you can **perform a rollback manually**.

- **IP address conflicts**

  Since the source and target servers have the same private IP addresses, there may be IP address conflicts. This may result in service unavailability.

- **Unknown risks**

  There may be other unknown issues since the migration does not detect or scan source services.

- **Customer responsibilities**

You need to fully test and prepare for the migration as well as check and solve possible problems after the migration is complete. You are advised to simulate the migration in a test environment to evaluate potential risks and formulate corresponding countermeasures.

Notes and Constraints

- **Shutting downing target servers**

  To retain source servers' private IP addresses on the paired target servers, the target servers must be stopped. If a target server is not stopped, the system will stop it automatically.
- **Subnet requirements**

  When you select a subnet in the target VPC, the subnet must be in the same network range as the source servers.
- **Network interface requirements**

  A target server can only have one network interface. Extended network interfaces are not allowed for target servers.
- **Supported IP version**

  Only IPv4 addresses can be retained.

Preparations

- Preparing a Huawei account

  Before using MgC, prepare a HUAWEI ID or an IAM user that can access MgC and obtain an AK/SK pair for the account or IAM user. For details, see **Preparations**.
- Creating a migration project

  Create a migration project on the MgC console. For details, see **Managing Migration Projects**.

Procedure

Step 1 Download and install the MgC Agent.

Prepare a Windows server on the source intranet for installing the MgC Agent. For details about the server requirements and the MgC Agent installation method, see **Installing the MgC Agent on Windows**.

Step 2 Connect the MgC Agent to MgC.

Log in to the MgC Agent console and connect the MgC Agent to MgC. For details, see **Connecting the MgC Agent to MgC**.

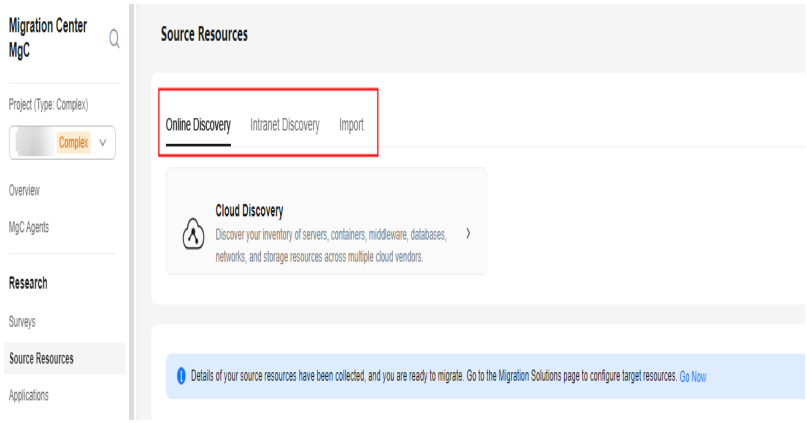
Step 3 Add resource credentials.

After the connection is successful, add the credentials of the source servers to be migrated to the MgC Agent. For details, see **Adding Resource Credentials**. Correctly configure **Resource Type** based on the source servers when you add their credentials.

Step 4 Discover source servers.

MgC provides three collection methods to meet your requirements in different scenarios. You can choose a method based on your source environment.

Figure 2-6 Collection methods



- If your source servers are on a cloud platform, such as Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, or Azure, you can **collect the data about your servers over the Internet**. You can also **manually add the server data to MgC**.
- If your source servers are in an on-premises IDC, you can **collect the server data over the intranet**. MgC enables you to scan for servers by network range or VMware vCenter Server. You can also **manually add the server data to MgC**.

Step 5 Group servers as an application.

Group the discovered servers as an application to get sizing recommendations and execute the migration. For more information, see **Grouping Resources as Applications**.

Step 6 Create an application assessment.

The system will generate target server recommendations based on the collected source server details, including specifications, performance data, and workload types. For more information, see **Getting Target Recommendations**.

After the target server recommendations are generated, you can:

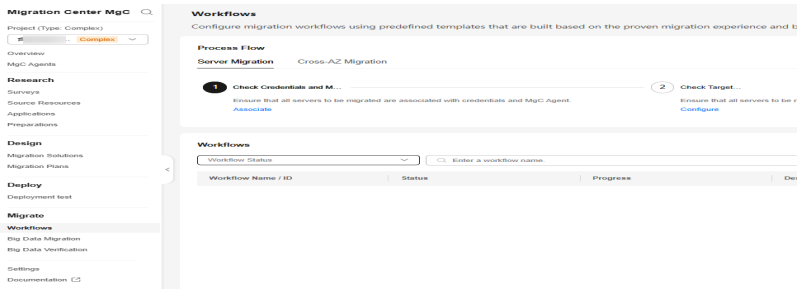
- **Modify the recommended target configurations** as needed. You can change the server and disk specifications. Disk downsizing is supported.
- **Associate source servers with target servers**. If you already have servers that match your requirements on Huawei Cloud, you can associate them with source servers.

Step 7 Create a server migration workflow. After all the preceding steps are complete, go to the **Workflows** page and create a server migration workflow.

CAUTION

- A workflow can contain a maximum of 100 servers.
- You can migrate a maximum of 1,000 servers concurrently. For any servers beyond this number, the workflows will pause at the first step and put these servers in a pending state until other servers complete their migration. The workflows will then automatically start on these servers in the order the workflows were created.
- If this is your first time to create a server migration workflow, you need to assign MgC the required permissions. For more information about the required permissions, see **Agency Permissions**.

1. In the navigation pane, choose **Workflows**. In the upper left corner of the page, select the **migration project you created**.
2. Select the **Server Migration** template and click **Configure Workflow**.



3. In the **Workflow Details** area, customize **Name** and **Description**.
4. In the **Application** area, select the application you created in **Step 5**.

NOTICE

For source servers in the selected applications, you need to **get target server recommendations** or **associate them with target servers**.

5. In the **Migration Network** area, select a network type.
   - If you select **Public**, ensure that all target servers have EIPs bound. These EIPs will be used for the migration.
   - If you select **Private**, configure Direct Connect connections, VPN connections, VPC peering connections, or subnets in the target VPC in advance to connect the source environment to the target environment.
6. In the **Target Environment** area, select the VPC and subnet that have been connected to the source environment. The VPC and subnet will be used as the transit environment.

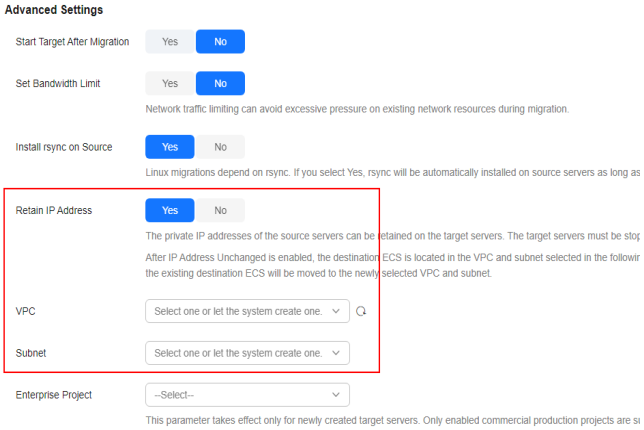
NOTICE

The configured VPC and subnet will not be applied for the target servers associated with source servers. Parameters in **Target Environment** are not required if all source servers in the selected application are associated with target servers.

7. In the **Advanced Settings** area, set **Retain IP Address** to **Yes**, read the on-screen warning and **disclaimer**, and click **Confirm**.

   NOTE

   Here only describes how to configure the **Retain IP Address** parameter. Set other parameters as required.



8. From the **VPC** drop-down list, select the VPC that contains the private network segment of the source servers. From the **Subnet** drop-down list, select the subnet that contains the private IP addresses of the source servers. After you select a subnet, the system will check whether the selected subnet contains the source servers' private IP addresses. If it does not, you need to change the subnet.
9. Click **Next: Confirm**.
10. Confirm the workflow settings and click **Confirm**. The **Run Workflow** dialog box is displayed, which indicates that the workflow has been created.
    - If you want to start the migration immediately, click **Confirm** to run the workflow.
    - If you want to **add a stage or step** to the workflow, click **Cancel**. The workflow enters a **Waiting** state, and the migration has not started. To start the migration, click **Run** in the **Operation** column.
11. On the migration workflow details page, view the workflow settings and the migration progress. After the step for starting the migration Agent is completed, a migration task is automatically created on the SMS console. For details about the server information mapping between MgC and SMS, see **What Are the Information Mappings Between MgC and SMS?**
    - Move the cursor to the migration progress bar. In the box that is displayed, view more migration details.
    - When the migration progress bar reaches a step that requires manual confirmation, move the cursor to the progress bar and click **Confirm** next to the step status in the displayed window, so that the subsequent migration steps can be executed.
    - When the workflow reaches the **ResizeDiskPartition** step, the system identifies whether disk capacity reduction has been performed on the target server.
      - If yes, go to **SMS console** and resize disks and partitions for the target server. For details, see the **Partition Resizing** parameter in

Configuring a Target Server. After the adjustment is complete, go back to the MgC console and click **Confirm** next to the step status so that the workflow can continue.

- If no, skip this step.

- The **StartSynchronization** step is repeated before you verify your services on the target server.
- When the progress bar reaches **Cutover**, the migration is complete. You need check whether your service systems are running properly on the target server. If they are, manually switch services to the target server. After the switchover is complete, click **Confirm** in the workflow. The system automatically performs the following steps **SourceClear** and **MigrationTaskClear**.

----End

Manual Rollback

If you do not want to retain the private IP addresses of the source servers or any problems happen, you can manually switch the VPC for rollback. For details, see **Changing a VPC**.

2.7 Batch Modifying and Restoring the Host Configurations for Linux Source Servers

2.7.1 Overview

Background

When you migrate a Linux source server, you need to ensure that the source server can identify and resolve the interface domain names of related cloud services. This usually involves editing the **hosts** file on the source server. If there are a large number of servers to be migrated, manual editing is time-consuming. To simplify this process, we provide example scripts for batch editing. You can use the scripts to quickly push the mappings between domain names and IP addresses to the **hosts** file on all source servers in batches.

Script Description

The scripts are developed using the Shell language and can run only on Linux. They are used to batch update and restore the **/etc/hosts** file on Linux source servers.

To prevent long script execution, a maximum of 100 servers can be modified at a time.

The following table describes for what and where the scripts are used.

Script	Function	Scenario
update_hosts_linux.sh	Batch update the **/etc/hosts** file on Linux source servers. The script will automatically log operations, alert for any exceptions, and generate a summary of the operations.	Before migrating Linux source servers, run this script to update the **hosts** file for the servers.
rollback_hosts_linux.sh	Batch restore the **/etc/hosts** file on the Linux source servers. The script will automatically log operations, alert for any exceptions, and generate a summary of the operations.	After the Linux source servers are migrated, run this script to restore the **hosts** file on the source servers to the state before the migration.

2.7.2 Preparations

Preparing a CSV File

Create a CSV file and write the source server information in the following format to file. Ensure that the file can be accessed.

```
username,ip,port,password
```

- **username**: indicates the username for logging in to the source server. To ensure that the scripts have sufficient permissions to perform the modification, you need to run them as a user with administrator permissions, such as **root**.
- **ip**: indicates the private IP address of the source server.
- **port**: indicates the listening port of the SSH service. By default, port 22 is used on Linux. If the SSH service of the source server is running on another port, specify the port correctly.
- **password**: indicates the password for logging in to the source server. The scripts use this password to automatically connect to the source server through SSH.

CAUTION

- The first line in the CSV file is the title line and will not be parsed by the scripts.
- Each line in the CSV file contains the information for a single server, with each piece of information separated by a comma.
- Ensure the format is correct and the information is accurate, avoiding any extra spaces, commas, or invalid IP addresses.

For example:


```
username,ip,port,password
root,192.168.1.10,xx,examplePass123
root,192.168.1.11,xx,examplePass456
```

Preparing the Hosts File

Create a text file that contains the content to be added to the **/etc/hosts** file on the source servers. Ensure that the file can be accessed. Write the API domain names mappings for the related cloud services into the file, and start with **#Migration-proxy-start** and end with **#Migration-proxy-end**. The API domain name mappings of related cloud services depend on the actual environment. Contact the environment contact person of the corresponding site to obtain the mappings.

For example:

```
#Migration-proxy-start
xxx.xxx.xxx.xxx  iam.xyz.com
xxx.xxx.xxx.xxx  ecs.xyz.com
xxx.xxx.xxx.xxx  evs.xyz.com
xxx.xxx.xxx.xxx  ims.xyz.com
xxx.xxx.xxx.xxx  obs.xyz.com
xxx.xxx.xxx.xxx  eps.xyz.com
xxx.xxx.xxx.xxx  vpc.xyz.com
#Migration-proxy-end
```

Preparing a Log Directory

- **Configure a log directory**. The scripts use **/var/log/update_hosts** as the default log storage directory. If the directory cannot be found, the scripts automatically create it. To change the log storage directory, change the value of **LOG_DIR** in the script.
- **Check the permissions for the log directory**. Ensure that the current user has the write permission for the log directory. If the permissions are insufficient, modify the directory permissions or use another directory.

Checking the Connectivity of the Source Servers

- **Check the network connection**. Ensure that the server where the scripts are executed can access all source servers over the network.
- **Check whether the SSH port is reachable**. Ensure that the SSH port (22 by default) on the source servers is reachable from the server where the scripts are executed.

Checking the SSH Configuration

- **Configure the SSH service**. Ensure that the SSH service has been enabled and is running properly on all source servers. The SSH service is enabled by default on most Linux distributions.
- **Enable SSH password authentication**. Ensure that the SSH service on all source servers is configured to accept password authentication so that the scripts can use the password provided in the CSV file for automatic login.

Configuring the Script Executor

Ensure that the following tools and commands are installed on the Linux server where the scripts are executed:

- SSH: used to establish secure connections to remote source servers.
- sed: used to edit and modify the **/etc/hosts** file.
- setsid: used to avoid interaction during SSH connections, typically to prevent prompts during password input.
- mktemp: used to create temporary files or directories.

Configuring the Execution and User Permissions

- **Configure user permissions**. Ensure that the user who executes the scripts has the read and write permissions for the log directory, CSV file, and **hosts** file.
- **Assign execute permission to the scripts**. Ensure that the **update_hosts_linux.sh** and **rollback_hosts_linux.sh** scripts are executable. Run the **chmod +x update_hosts_linux.sh** and **chmod +x rollback_hosts_linux.sh** commands to add the execute permission to the scripts.

2.7.3 Configuring the Scripts

2.7.3.1 Configuring the update_hosts_linux.sh Script

Modify the configuration in the example script to meet your specific requirements.

Prerequisites

You have completed all **preparations**.

Procedure

Step 1 Create a file named **update_hosts_linux.sh** on the server where the script is executed, and copy the following script content to the file. If you have connected to the Linux source servers through SSH, you can directly use Vim to create and edit a script file. The procedure is as follows:

1. In the Vim editor, press **I** to enter insert mode.
2. Copy and paste the following script code and press **Esc**.
3. Run **:wq** to save and exit.

```
#!/bin/bash

# Configuration
# Log directory path: Used to store run logs, error logs, and summary logs.
# If the directory doesn't exist, the script will create it automatically.
LOG_DIR="/var/log/update_hosts"

# Run log file path: Records detailed information about the script's execution.
RUN_LOG="$LOG_DIR/run.log"

# Error log file path: Records any errors that occur during the script's execution.
ERROR_LOG="$LOG_DIR/error.log"

# Summary log file path: Records a summary of the script's execution, including the number of successful and failed servers.
SUMMARY_LOG="$LOG_DIR/summary.log"

# CSV file path: Contains information about the target hosts (must be manually created and configured).
```



```
CSV_FILE="target_servers.csv"

# Hosts content file path: Contains the content to be appended to each target host's /etc/hosts file (must be manually created and configured).
HOSTS_FILE="hosts_content.txt"
DEFAULT_PORT=22
SSH_TIMEOUT=10

# Initialize log directory and files
initialize_logs() {
    mkdir -p "$LOG_DIR"
    echo "========================================" >> "$RUN_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new update execution" >> "$RUN_LOG"
    echo "========================================" >> "$RUN_LOG"

    echo "========================================" >> "$ERROR_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new update execution" >> "$ERROR_LOG"
    echo "========================================" >> "$ERROR_LOG"

    echo "========================================" > "$SUMMARY_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new update execution" >> "$SUMMARY_LOG"
    echo "========================================" >> "$SUMMARY_LOG"
}

# Log info function
log_info() {
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG"
}

# Log error function
log_error() {
    echo "[ERROR] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG" "$ERROR_LOG"
}

# Read server information from CSV file
read_servers_from_csv() {
    local csv_file="$1"
    local servers=()
    local header_skipped=false

    if [ ! -f "$csv_file" ]; then
        log_error "CSV file '$csv_file' not found."
        exit 1
    fi

    # Ensure file ends with a newline character
    sed -i -e '$a\' "$csv_file"

    while IFS=, read -r username ip port password; do
        # Skip header row
        if [ "$header_skipped" = false ]; then
            header_skipped=true
            continue
        fi
        # Skip empty and invalid rows
        if [[ -z "$username" || -z "$ip" ]]; then
            continue
        fi
        port=${port:-$DEFAULT_PORT} # Use default port 22
        # Ensure port is numeric
        if ! [[ "$port" =~ ^[0-9]+$ ]]; then
            log_error "Invalid port '$port' for $username@$ip. Skipping this server."
            continue
        fi
        servers+=("$username@$ip:$port:$password")
    done < "$csv_file"

    echo "${servers[@]}"
}
```



```
# Read hosts content from TXT file
read_hosts_content_from_txt() {
    local txt_file="$1"
    if [ -f "$txt_file" ]; then
        cat "$txt_file"
    else
        log_error "Hosts content file '$txt_file' not found."
        exit 1
    fi
}

# Initialize log files
initialize_logs

# Read server information from CSV file
servers=($(read_servers_from_csv "$CSV_FILE"))

# Read hosts content from TXT file
hosts_content=$(read_hosts_content_from_txt "$HOSTS_FILE")

# Counters for success and failure
success_count=0
failure_count=0
failed_servers=()

# Iterate over each server and push hosts content
for server in "${servers[@]}"; do
  # Extract user, ip, port, and password information
  IFS=':' read -r user_host port pass <<< "$server"
  IFS='@' read -r user ip <<< "$user_host"

  log_info "Starting update for $user@$ip:$port"

  # Create temporary script and SSH_ASKPASS script
  tmp_script=$(mktemp)
  askpass_script=$(mktemp)

  cat <<EOF > "$tmp_script"
#!/bin/bash
# Backup hosts file
if [ ! -f /etc/hosts.bak ]; then
  cp /etc/hosts /etc/hosts.bak
fi

# Remove old Migration-proxy section
sed -i '/#Migration-proxy-start/,/#Migration-proxy-end/d' /etc/hosts

# Append new Migration-proxy section
echo "$hosts_content" >> /etc/hosts
EOF

  cat <<EOF > "$askpass_script"
#!/bin/bash
echo "$pass"
EOF

  chmod +x "$tmp_script" "$askpass_script"

  # Set SSH_ASKPASS environment variable and use ssh to connect to the target machine and execute the
temporary script
  export SSH_ASKPASS="$askpass_script"
  export DISPLAY=:0

  ssh_output=$(mktemp)
  setsid ssh -o BatchMode=no -o ConnectTimeout=$SSH_TIMEOUT -o StrictHostKeyChecking=no -p "$port"
"$user@$ip" 'bash -s' < "$tmp_script" 2> "$ssh_output"
  ssh_status=$?
```



```
  if [ $ssh_status -eq 0 ]; then
    log_info "Updated hosts on $ip:$port successfully"
    ((success_count++))
  else
    ssh_error=$(cat "$ssh_output")
    case $ssh_status in
      1)
        log_error "General error occurred while updating hosts on $ip:$port: $ssh_error"
        ;;
      2)
        log_error "Misuse of shell builtins while updating hosts on $ip:$port: $ssh_error"
        ;;
      255)
        if [[ "$ssh_error" == *"Permission denied"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Permission denied (password may be incorrect or
username is wrong)"
        elif [[ "$ssh_error" == *"Connection refused"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Connection refused (port may be incorrect or SSH
service not running on target)"
        elif [[ "$ssh_error" == *"No route to host"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: No route to host (network unreachable)"
        elif [[ "$ssh_error" == *"Host key verification failed"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Host key verification failed"
        elif [[ "$ssh_error" == *"Connection timed out"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Connection timed out"
        else
          log_error "SSH login failed for $user@$ip:$port: $ssh_error"
        fi
        ;;
      *)
        log_error "An unknown error occurred while updating hosts on $ip:$port: $ssh_error"
        ;;
    esac
    failed_servers+=("$user@$ip:$port")
    ((failure_count++))
  fi

  # Remove temporary scripts and SSH output file
  rm -f "$tmp_script" "$askpass_script" "$ssh_output"
done

# Calculate failure and success percentages
total_count=${#servers[@]}
failure_percentage=$(echo "scale=2; ($failure_count / $total_count) * 100" | bc)
success_percentage=$(echo "scale=2; ($success_count / $total_count) * 100" | bc)

# Output summary result and log to file
summary_content=$(cat <<EOF
========================================
[SUMMARY] $(date '+%Y-%m-%d %H:%M:%S') - Execution Update Summary
========================================
Total number of servers: $total_count
Number of successful updates: $success_count
Number of failed updates: $failure_count
Success rate: $success_percentage%
Failure rate: $failure_percentage%
----------------------------------------
EOF
)

if [ $failure_count -gt 0 ]; then
    summary_content+="Failed servers:\n"
    for server in "${failed_servers[@]}"; do
        summary_content+="  - $server\n"
    done
fi
summary_content+="========================================"

# Output summary result to log file and terminal
```



```
echo -e "$summary_content" | tee -a "$SUMMARY_LOG"

log_info "script execution completed. Check $SUMMARY_LOG for summary."
```

Step 2 Modify the following parameters in the script to meet your needs:

- **LOG_DIR="/var/log/update_hosts"**
  - Description: log directory, which is used to store run, error, and summary logs.
  - Default value: **/var/log/update_hosts**
  - Suggestion: Change the value to a directory for which the current user has the write permission.
  - Example: **LOG_DIR="/home/username/update_hosts_logs"**
- **CSV_FILE="target_servers.csv"**
  - Description: **CSV file** path. The file contains the source server information.
  - Default value: **target_servers.csv**
  - Suggestion: Use an absolute path or a correct relative path.
  - Example: **CSV_FILE="/home/username/configs/servers.csv"**
- **HOSTS_FILE="hosts_content.txt"**
  - Description: **hosts** file path. The file contains the content to be added to the **/etc/hosts** file on the source servers.
  - Default value: **hosts_content.txt**
  - Suggestion: Use an absolute path or a correct relative path.
  - Example: **HOSTS_FILE="/home/username/configs/hosts_content.txt"**

Step 3 Save the changes. Run the script in a terminal window. If a GUI is available, press **Ctrl**+**Alt**+**T** to open the terminal.

```
./update_hosts_linux.sh
```

The script will output log information in the terminal window and generate a result report upon completion. You can view this report in the **summary.log** file in the directory specified by **LOG_DIR**.

----End

2.7.3.2 Configuring the rollback_hosts_linux.sh Script

Modify the configuration in the example script to meet your specific requirements.

Prerequisites

You have completed all **preparations**.

Procedure

Step 1 Create a file named **rollback_hosts_linux.sh** on the server where the script is executed, and copy the following script content to the file. If you have connected to the Linux source servers through SSH, you can directly use Vim to create and edit a script file. The procedure is as follows:

1. In the Vim editor, press **I** to enter insert mode.
2. Copy and paste the following script code and press **Esc**.
3. Run **:wq** to save and exit.

```
#!/bin/bash

# Configuration

# Log directory path: Used to store run logs, error logs, and summary logs.
# If the directory doesn't exist, the script will create it automatically.
LOG_DIR="/var/log/update_hosts"

# Run log file path: Records detailed information about the script's execution.
RUN_LOG="$LOG_DIR/run.log"

# Error log file path: Records any errors that occur during the script's execution.
ERROR_LOG="$LOG_DIR/error.log"

# Summary log file path: Records a summary of the script's execution, including the number of successful and failed servers.
SUMMARY_LOG="$LOG_DIR/summary.log"

# CSV file path: Contains information about the target hosts (must be manually created and configured).
CSV_FILE="target_servers.csv"
DEFAULT_PORT=22
SSH_TIMEOUT=10

# Initialize log directory and files
initialize_logs() {
    mkdir -p "$LOG_DIR"
    echo "========================================" >> "$RUN_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new rollback execution" >> "$RUN_LOG"
    echo "========================================" >> "$RUN_LOG"

    echo "========================================" >> "$ERROR_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new rollback execution" >> "$ERROR_LOG"
    echo "========================================" >> "$ERROR_LOG"

    echo "========================================" > "$SUMMARY_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new rollback execution" >> "$SUMMARY_LOG"
    echo "========================================" >> "$SUMMARY_LOG"
}

# Log info function
log_info() {
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG"
}

# Log error function
log_error() {
    echo "[ERROR] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG" "$ERROR_LOG"
}

# Read server information from CSV file
read_servers_from_csv() {
    local csv_file="$1"
    local servers=()
    local header_skipped=false

    if [ ! -f "$csv_file" ]; then
        log_error "CSV file '$csv_file' not found."
        exit 1
    fi

    # Ensure file ends with a newline character
    sed -i -e '$a\' "$csv_file"

    while IFS=, read -r username ip port password; do
```

```
        # Skip header row
        if [ "$header_skipped" = false ]; then
            header_skipped=true
            continue
        fi
        # Skip empty and invalid rows
        if [[ -z "$username" || -z "$ip" ]]; then
            continue
        fi
        port=${port:-$DEFAULT_PORT} # Use default port 22
        # Ensure port is numeric
        if ! [[ "$port" =~ ^[0-9]+$ ]]; then
            log_error "Invalid port '$port' for $username@$ip. Skipping this server."
            continue
        fi
        servers+=("$username@$ip:$port:$password")
    done < "$csv_file"

    echo "${servers[@]}"
}

# Initialize log files
initialize_logs

# Read server information from CSV file
servers=($(read_servers_from_csv "$CSV_FILE"))

# Counters for success and failure
success_count=0
failure_count=0
failed_servers=()

# Iterate over each server and execute rollback
for server in "${servers[@]}"; do
  # Extract user, ip, port, and password information
  IFS=':' read -r user_host port pass <<< "$server"
  IFS='@' read -r user ip <<< "$user_host"

  log_info "Starting rollback for $user@$ip:$port"

  # Create temporary script and SSH_ASKPASS script
  tmp_script=$(mktemp)
  askpass_script=$(mktemp)

  cat <<EOF > "$tmp_script"
#!/bin/bash
# Backup hosts file
if [ ! -f /etc/hosts.bak ]; then
  cp /etc/hosts /etc/hosts.bak
fi

# Remove old Migration-proxy section
sed -i '/#Migration-proxy-start/,/#Migration-proxy-end/d' /etc/hosts
EOF

  cat <<EOF > "$askpass_script"
#!/bin/bash
echo "$pass"
EOF

  chmod +x "$tmp_script" "$askpass_script"

  # Set SSH_ASKPASS environment variable and use ssh to connect to the target machine and execute the
temporary script
  export SSH_ASKPASS="$askpass_script"
  export DISPLAY=:0

  ssh_output=$(mktemp)
  setsid ssh -o BatchMode=no -o ConnectTimeout=$SSH_TIMEOUT -o StrictHostKeyChecking=no -p "$port"
```



```
"$user@$ip" 'bash -s' < "$tmp_script" 2> "$ssh_output"
  ssh_status=$?

  if [ $ssh_status -eq 0 ]; then
    log_info "Rolled back hosts on $ip:$port successfully"
    ((success_count++))
  else
    ssh_error=$(cat "$ssh_output")
    case $ssh_status in
      1)
        log_error "General error occurred while rolling back hosts on $ip:$port: $ssh_error"
        ;;
      2)
        log_error "Misuse of shell builtins while rolling back hosts on $ip:$port: $ssh_error"
        ;;
      255)
        if [[ "$ssh_error" == *"Permission denied"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Permission denied (password may be incorrect or
username is wrong)"
        elif [[ "$ssh_error" == *"Connection refused"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Connection refused (port may be incorrect or SSH
service not running on target)"
        elif [[ "$ssh_error" == *"No route to host"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: No route to host (network unreachable)"
        elif [[ "$ssh_error" == *"Host key verification failed"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Host key verification failed"
        elif [[ "$ssh_error" == *"Connection timed out"* ]]; then
          log_error "SSH login failed for $user@$ip:$port: Connection timed out"
        else
          log_error "SSH login failed for $user@$ip:$port: $ssh_error"
        fi
        ;;
      *)
        log_error "An unknown error occurred while rolling back hosts on $ip:$port: $ssh_error"
        ;;
    esac
    failed_servers+=("$user@$ip:$port")
    ((failure_count++))
  fi

  # Remove temporary scripts and SSH output file
  rm -f "$tmp_script" "$askpass_script" "$ssh_output"
done

# Calculate failure and success percentages
total_count=${#servers[@]}
failure_percentage=$(echo "scale=2; ($failure_count / $total_count) * 100" | bc)
success_percentage=$(echo "scale=2; ($success_count / $total_count) * 100" | bc)

# Output summary result and log to file
summary_content=$(cat <<EOF
========================================
[SUMMARY] $(date '+%Y-%m-%d %H:%M:%S') - Execution Rollback Summary
========================================
Total number of servers: $total_count
Number of successful rollbacks: $success_count
Number of failed rollbacks: $failure_count
Success rate: $success_percentage%
Failure rate: $failure_percentage%
----------------------------------------
EOF
)

if [ $failure_count -gt 0 ]; then
    summary_content+="Failed servers:\n"
    for server in "${failed_servers[@]}"; do
        summary_content+="  - $server\n"
    done
fi
```



```
summary_content+="======================================="

# Output summary result to log file and terminal
echo -e "$summary_content" | tee -a "$SUMMARY_LOG"

log_info "script execution completed. Check $SUMMARY_LOG for summary."
```

Step 2 Modify the following parameters in the script to meet your needs:

- **LOG_DIR="/var/log/rollback_hosts"**
  - Description: log directory
  - Default value: **/var/log/rollback_hosts**
  - **Suggestion**: Change the value to a directory for which the current user has the write permission.
  - Example: **LOG_DIR="/home/username/rollback_hosts_logs"**
- **CSV_FILE="target_servers.csv"**
  - Description: CSV file path. The file contains the source server information.
  - Default value: **target_servers.csv**
  - Suggestion: Use an absolute path or a correct relative path.
  - Example: **CSV_FILE="/home/username/configs/servers.csv"**

Step 3 Save the changes. Run the script in a terminal window. If a GUI is available, press **Ctrl**+**Alt**+**T** to open the terminal.

```
./rollback_hosts_linux.sh
```

The script will output log information in the terminal window and generate a result report upon completion. You can view this report in the **summary.log** file in the directory specified by **LOG_DIR**.

----End

2.8 Batch Modifying and Restoring the Host Configurations for Windows Source Servers

2.8.1 Overview

Background

When you migrate a Windows source server, you need to ensure that the source server can resolve the interface domain names of related cloud services. This usually involves editing the **hosts** file on the source server. If there are a large number of servers to be migrated, manual editing is time-consuming. To simplify this process, we provide example scripts for batch editing. You can use the scripts to quickly write the mappings between domain names and IP addresses to the **hosts** file on all source servers in batches.

Script Description

The scripts are developed using the PowerShell language and can run only on Windows. They are used to batch update and restore the **/etc/hosts** file on

Windows source servers. The path of the **hosts** file is **C:\Windows \System32\drivers\etc\hosts**.

To prevent long script execution, a maximum of 100 servers can be modified at a time.

The following table describes for what and where the scripts are used.

Script	Description	Scenario
update_hosts_win.ps1	Batch update the **hosts** file on Windows source servers. The script will automatically log operations, alert for any exceptions, and generate a summary of the operations.	Before migrating Windows source servers, run this script to update the **hosts** file for the servers.
rollback_hosts_win.ps1	Batch restore the **hosts** file on Windows source servers. The script will automatically log operations, alert for any exceptions, and generate a summary of the operations.	After the Windows source servers are migrated, run this script to restore the **hosts** file on these servers to the state before the migration.

PowerShell Version and Dependency Requirements

PowerShell remoting uses WinRM, which provided by the Windows Management Framework (WMF).

To run remote sessions on PowerShell, the local and remote computers must have the following:

- Windows PowerShell 3.0 or later (WMF 5.1 is recommended.)
- Microsoft .NET Framework 4.0 or later
- WinRM 3.0 or later

To run remote sessions on Windows PowerShell 2.0, the local and remote computers must have the following:

- Windows PowerShell 2.0 or later
- Microsoft .NET Framework 2.0 or later
- WinRM 2.0

Features that run only on Windows PowerShell 3.0 or higher, such as the ability to disconnect and reconnect to sessions, are only available when both computers are running Windows PowerShell 3.0 or higher.

Run the following command to check the PowerShell version:

```
$PSVersionTable
```


2.8.2 Preparations

Configuring a PowerShell Execution Policy

Check the execution policy and ensure that PowerShell allows script execution. Open PowerShell and run the following command to check the current execution policy:

```
Get-ExecutionPolicy
```

The PowerShell execution policies are as follows:

- Restricted: No script can be executed.
- AllSigned: Only scripts signed by trusted publishers can be run.
- RemoteSigned: Locally created scripts can be run without signatures, but remotely downloaded scripts must be signed.
- Unrestricted: All scripts can be executed, but a warning is generated when a script downloaded from the Internet is executed.
- Bypass: Nothing is blocked and there are no warnings or prompts.

If the execution policy is **Restricted** or **AllSigned**, run the following command to temporarily change the policy to allow the execution of locally created scripts and signed remote scripts:

```
Set-ExecutionPolicy RemoteSigned -Scope Process
```

This command changes the execution policy only in the current PowerShell session and restores the default policy after the session ends.

Preparing a CSV File

Create a CSV file and write the source server information in the following format to file. Ensure that the file can be accessed.

```
username,ip,port,password
```

- **username**: indicates the username for logging in to the source server. To ensure that the script has sufficient permissions to perform the modification, you need to run it as a user with administrator permissions, such as **Administrator**.
- **ip**: indicates the private IP address of the source server.
- **port**: listening port of the WinRM service. The default port is 5985.
- **password**: indicates the password for logging in to the source server. The scripts use this password to automatically connect to the source server through WinRM.

CAUTION

- The first row in the CSV file is the header row.
- Each line in the CSV file contains the information for a single server, with each piece of information separated by a comma.
- Ensure the format is correct and the information is accurate, avoiding any extra spaces, commas, or invalid IP addresses.

For example:

```
username,ip,port,password
Administrator,192.168.1.10,xx,examplePass123
Administrator,192.168.1.11,xx,examplePass456
```

Preparing the Hosts File

Create a text file that contains the content to be added to the **hosts** file on the source servers. Ensure that the file can be accessed. Write the API domain names mappings for the related cloud services into the file, and start with **#Migration-proxy-start** and end with **#Migration-proxy-end**. The API domain name mappings of related cloud services depend on the actual environment. Contact the environment contact person of the corresponding site to obtain the mappings.

For example:

```
#Migration-proxy-start
xxx.xxx.xxx.xxx  iam.xyz.com
xxx.xxx.xxx.xxx  ecs.xyz.com
xxx.xxx.xxx.xxx  evs.xyz.com
xxx.xxx.xxx.xxx  ims.xyz.com
xxx.xxx.xxx.xxx  obs.xyz.com
xxx.xxx.xxx.xxx  eps.xyz.com
xxx.xxx.xxx.xxx  vpc.xyz.com
#Migration-proxy-end
```

Preparing a Log Directory

- **Configure the log directory**. The scripts use **C:\Users\Public\Hosts_script_Logs** as the default log storage directory. If the directory cannot be found, the scripts automatically create it. To change the log storage directory, change the value of **$logDir** in the script.
- **Check the permissions for the log directory**. Ensure that the current user has the write permission for the log directory. If the permissions are insufficient, modify the directory permissions or use another directory.

Checking the Network Connectivity

- **Check the network connection**. Ensure that the server where the scripts are executed can access the IP addresses and ports of all Windows source servers over the network. The script executor must be able to access all Windows source servers over port 5985.
- **Configure the firewalls**. Check and configure the firewalls on the local computer and source servers to ensure that remote PowerShell sessions can be established through WinRM.
- **Enable the WinRM service**. Ensure that the WinRM service has been enabled and is running properly on all Windows source servers. You can run the following command on the source servers to enable WinRM:

  ```
  Enable-PSRemoting -Force
  ```

2.8.3 Example Scripts

2.8.3.1 Configuring the update_hosts_win.ps1 Script

Modify the configuration in the example script to meet your specific requirements.

Prerequisites

You have completed all **preparations**.

Procedure

Step 1 Create a file named **update_hosts_lwin.ps1** on the server where the script is executed, and copy the following script content to the file.

```
# Configuration
# Path to the CSV file with server information. Must exist before running the script.
$csvFile = "C:\Users\Public\target_servers.csv"  # Manually configure

# Path to the hosts content file. Must exist before running the script.
$hostsFile = "C:\Users\Public\hosts_content.txt"  # Manually configure

# Directory for storing log files. Will be created if it doesn't exist.
$logDir = "C:\Users\Public\Hosts_script_Logs"  # Automatically created

# Log file for general run information.
$runLog = Join-Path $logDir "run.log"  # Automatically created

# Log file for error messages.
$errorLog = Join-Path $logDir "error.log"  # Automatically created

# Log file for summary information.
$summaryLog = Join-Path $logDir "summary.log"  # Automatically created


# Initialize log directory and files
function Initialize-Logs {
    if (-not (Test-Path $logDir)) {
        New-Item -Path $logDir -ItemType Directory
    }
    Add-Content -Path $runLog -Value "========================================"
    Add-Content -Path $runLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting new
update execution"
    Add-Content -Path $runLog -Value "========================================"

    Add-Content -Path $errorLog -Value "========================================"
    Add-Content -Path $errorLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting
new update execution"
    Add-Content -Path $errorLog -Value "========================================"

    Add-Content -Path $summaryLog -Value "========================================"
    Add-Content -Path $summaryLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') -
Starting new update execution"
    Add-Content -Path $summaryLog -Value "========================================"
}

# Log info function
function Log-Info {
    param (
        [string]$message
    )
    $logMessage = "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Write-Output $logMessage
}

# Log error function
function Log-Error {
    param (
        [string]$message
    )
    $logMessage = "[ERROR] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Add-Content -Path $errorLog -Value $logMessage
```



```
    Write-Output $logMessage
}

# Read server information from CSV file
function Read-ServersFromCSV {
    param (
        [string]$csvFile
    )
    if (-not (Test-Path $csvFile)) {
        Log-Error "CSV file '$csvFile' not found."
        exit 1
    }
    return Import-Csv -Path $csvFile
}

# Read hosts content from TXT file
function Read-HostsContentFromTXT {
    param (
        [string]$hostsFile
    )
    if (-not (Test-Path $hostsFile)) {
        Log-Error "Hosts content file '$hostsFile' not found."
        exit 1
    }
    return Get-Content -Path $hostsFile -Raw
}

# Add to TrustedHosts
function Add-ToTrustedHosts {
    param (
        [string]$ip
    )
    # Check current TrustedHosts list
    $trustedHostsPath = "WSMan:\localhost\Client\TrustedHosts"
    $trustedHosts = (Get-Item $trustedHostsPath).Value

    if ($trustedHosts -eq $null -or $trustedHosts -eq "") {
        # Set the initial trusted host
        Set-Item $trustedHostsPath -Value $ip -Force
        Log-Info "Set initial TrustedHosts value to $ip"
    } elseif ($trustedHosts -notlike "*$ip*") {
        # Add new IP to TrustedHosts if not already present
        $updatedTrustedHosts = if ($trustedHosts -eq "*") { $ip } else { "$trustedHosts,$ip" }
        try {
            Set-Item $trustedHostsPath -Value $updatedTrustedHosts -Force
            Log-Info "Added $ip to TrustedHosts"
        } catch {
            Log-Error "Failed to add $ip to TrustedHosts: $_"
        }
    } else {
        Write-Host "TrustedHosts list already contains IP $ip."
    }
}

# Initialize log files
Initialize-Logs

# Verify CSV file
if (-not (Test-Path $csvFile)) {
    Log-Error "CSV file '$csvFile' not found."
    exit 1
}

# Verify hosts file
if (-not (Test-Path $hostsFile)) {
    Log-Error "Hosts content file '$hostsFile' not found."
    exit 1
}
```



```
# Read server information from CSV file
$servers = Read-ServersFromCSV -csvFile $csvFile

# Read hosts content from TXT file
$hostsContent = Read-HostsContentFromTXT -hostsFile $hostsFile

# Counters for success and failure
$successCount = 0
$failureCount = 0
$failedServers = @()

# Remote script block
$remoteScriptBlock = {
    param (
        [string]$hostsContent
    )
    $hostsFilePath = 'C:\Windows\System32\drivers\etc\hosts'

    # Read the file content
    $content = Get-Content -Path $hostsFilePath

    # Initialize flag
    $inBlock = $false
    $newContent = @()

    # Traverse file content
    foreach ($line in $content) {
        if ($line -match '#Migration-proxy-start') {
            $inBlock = $true
        }
        if (-not $inBlock) {
            $newContent += $line
        }
        if ($line -match '#Migration-proxy-end') {
            $inBlock = $false
            continue
        }
    }

    # Remove trailing empty lines
    while ($newContent[-1] -eq '') {
        $newContent = $newContent[0..($newContent.Count - 2)]
    }

    # Write the new content back to the file
    $newContent | Set-Content -Path $hostsFilePath

    # Append new Migration-proxy section
    Add-Content -Path $hostsFilePath -Value $hostsContent

    Write-Output 'Successfully updated hosts file on remote server.'
}

# Main script logic
Log-Info "Script execution started."

foreach ($server in $servers) {
    $username = $server.username
    $ip = $server.ip
    $password = $server.password

    if (-not $username -or -not ${ip} -or -not $password) {
        Log-Error "Invalid server entry: $username, ${ip}, $password. Skipping."
        continue
    }

    Log-Info "Starting update for $username@${ip}"

    $securePassword = ConvertTo-SecureString $password -AsPlainText -Force
```



```
    $credential = New-Object System.Management.Automation.PSCredential ($username, $securePassword)

    Add-ToTrustedHosts -ip $ip

    try {
        $session = New-PSSession -ComputerName ${ip} -Credential $credential -ErrorAction Stop

        Invoke-Command -Session $session -ScriptBlock $remoteScriptBlock -ArgumentList $hostsContent

        Remove-PSSession -Session $session
        Log-Info "Updated hosts on ${ip} successfully"
        $successCount++
    }
    catch {
        Log-Error "Failed to update hosts on ${ip}: $_"
        $failedServers += "$username@${ip}"
        $failureCount++
    }
}

# Calculate failure and success percentages
$totalCount = $servers.Count
if ($totalCount -gt 0) {
    $failurePercentage = [math]::Round(($failureCount / $totalCount) * 100, 2)
    $successPercentage = [math]::Round(($successCount / $totalCount) * 100, 2)
} else {
    $failurePercentage = 0
    $successPercentage = 100
}

# Output summary result and log to file
$summaryContent = @"
========================================
[SUMMARY] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Execution Update Summary
========================================
Total number of servers: $totalCount
Number of successful updates: $successCount
Number of failed updates: $failureCount
Success rate: $successPercentage%
Failure rate: $failurePercentage%
----------------------------------------
"@

if ($failedServers.Count -gt 0) {
    $summaryContent += "Failed servers:`n"
    foreach ($server in $failedServers) {
        $summaryContent += "  - $server`n"
    }
}
$summaryContent += "========================================"

# Output summary result to log file and terminal
$summaryContent | Add-Content -Path $summaryLog
Write-Output $summaryContent
Log-Info "Script execution completed. Check $summaryLog for summary."
```

Step 2 Modify the following parameters in the script to meet your needs:

- **$logDir = "C:\Users\Public\Hosts_Script_Logs"**
  - Description: log directory, which is used to store run, error, and summary logs.
  - Default value: **C:\Users\Public\Hosts_Script_Logs**
  - Suggestion: Change the value to a directory for which the current user has the write permission.
  - Example: **$logDir ="C:\Users\username\Documents\Hosts_Script_Logs"**

- **$csvFile = "C:\Users\Public\target_servers.csv"**
  - Description: **CSV file** path. The file contains the source server information.
  - Default value: **C:\Users\Public\target_servers.csv**
  - Suggestion: Use an absolute path or a correct relative path. If the CSV file path changes, you need to update the path here.
  - Example: **$csvFile = "C:\Users\username\Documents\servers.csv"**
- **$hostsFile = "C:\Users\Public\hosts_content.txt"**
  - Description: **hosts** file path. The file contains the content to be added to the **hosts** file on the source servers.
  - Default value: **C:\Users\Public\hosts_content.txt**
  - Suggestion: Use an absolute path or a correct relative path.
  - Example: **$hostsFile = "C:\Users\username\Documents\hosts_content.txt"**

Step 3 After the configuration items are modified and saved, run PowerShell as administrator and execute the script.

```
.\update_hosts_win.ps1
```

The script will output log information in the terminal window and generate a result report upon completion. You can view this report in the **summary.log** file in the directory specified by **$logDir**.

----End

2.8.3.2 Configuring the rollback_hosts_win.ps1 Script

Modify the configuration in the example script to meet your specific requirements.

Prerequisites

You have completed all **preparations**.

Procedure

Step 1 Create a file named **rollback_hosts_win.ps1** on the server where the script is executed, and copy the following script content to the file.

```
# Configuration
# Path to the CSV file with server information. Must exist before running the script.
$csvFile = "C:\Users\Public\target_servers.csv"  # Manually configure

# Directory for storing log files. Will be created if it doesn't exist.
$logDir = "C:\Users\Public\Hosts_Script_Logs"  # Automatically created

# Log file for general run information.
$runLog = Join-Path $logDir "run.log"  # Automatically created

# Log file for error messages.
$errorLog = Join-Path $logDir "error.log"  # Automatically created

# Log file for summary information.
$summaryLog = Join-Path $logDir "summary.log"  # Automatically created


# Initialize log directory and files
```



```
function Initialize-Logs {
    if (-not (Test-Path $logDir)) {
        New-Item -Path $logDir -ItemType Directory
    }
    Add-Content -Path $runLog -Value "========================================="
    Add-Content -Path $runLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting new
restore execution"
    Add-Content -Path $runLog -Value "========================================="

    Add-Content -Path $errorLog -Value "========================================="
    Add-Content -Path $errorLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting
new restore execution"
    Add-Content -Path $errorLog -Value "========================================="

    Add-Content -Path $summaryLog -Value "========================================="
    Add-Content -Path $summaryLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') -
Starting new restore execution"
    Add-Content -Path $summaryLog -Value "========================================="
}

# Log info function
function Log-Info {
    param (
        [string]$message
    )
    $logMessage = "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Write-Output $logMessage
}

# Log error function
function Log-Error {
    param (
        [string]$message
    )
    $logMessage = "[ERROR] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Add-Content -Path $errorLog -Value $logMessage
    Write-Output $logMessage
}

# Read server information from CSV file
function Read-ServersFromCSV {
    param (
        [string]$csvFile
    )
    if (-not (Test-Path $csvFile)) {
        Log-Error "CSV file '$csvFile' not found."
        exit 1
    }
    return Import-Csv -Path $csvFile
}

# Add to TrustedHosts
function Add-ToTrustedHosts {
    param (
        [string]$ip
    )
    # Check current TrustedHosts list
    $trustedHostsPath = "WSMan:\localhost\Client\TrustedHosts"
    $trustedHosts = (Get-Item $trustedHostsPath).Value

    if ($trustedHosts -eq $null -or $trustedHosts -eq "") {
        # Set the initial trusted host
        Set-Item $trustedHostsPath -Value $IP -Force
        Log-Info "Set initial TrustedHosts value to $IP"
    } elseif ($trustedHosts -notlike "*$ip*") {
        # Add new IP to TrustedHosts if not already present
        $updatedTrustedHosts = if ($trustedHosts -eq "*") { $ip } else { "$trustedHosts,$ip" }
```



```
        try {
            Set-Item $trustedHostsPath -Value $updatedTrustedHosts -Force
            Log-Info "Added $ip to TrustedHosts"
        } catch {
            Log-Error "Failed to add $ip to TrustedHosts: $_"
        }
    } else {
        Write-Host "TrustedHosts list already contains IP $ip."
    }
}

# Initialize log files
Initialize-Logs

# Verify CSV file
if (-not (Test-Path $csvFile)) {
    Log-Error "CSV file '$csvFile' not found."
    exit 1
}

# Read server information from CSV file
$servers = Read-ServersFromCSV -csvFile $csvFile

# Counters for success and failure
$successCount = 0
$failureCount = 0
$failedServers = @()

# Remote script block
$remoteScriptBlock = {
    param ()

    $hostsFilePath = 'C:\Windows\System32\drivers\etc\hosts'

    # Read the file content
    $content = Get-Content -Path $hostsFilePath

    # Initialize flag
    $inBlock = $false
    $newContent = @()

    # Traverse file content
    foreach ($line in $content) {
        if ($line -match '#Migration-proxy-start') {
            $inBlock = $true
        }
        if (-not $inBlock) {
            $newContent += $line
        }
        if ($line -match '#Migration-proxy-end') {
            $inBlock = $false
            continue
        }
    }

    # Remove trailing empty lines
    while ($newContent[-1] -eq '') {
        $newContent = $newContent[0..($newContent.Count - 2)]
    }

    # Write the new content back to the file
    $newContent | Set-Content -Path $hostsFilePath

    Write-Output 'Successfully restored hosts file on remote server.'
}

# Main script logic
Log-Info "Script execution started."
```



```
foreach ($server in $servers) {
    $username = $server.username
    $ip = $server.ip
    $password = $server.password

    if (-not $username -or -not ${ip} -or -not $password) {
        Log-Error "Invalid server entry: $username, ${ip}, $password. Skipping."
        continue
    }

    Log-Info "Starting restore for $username@${ip}"

    $securePassword = ConvertTo-SecureString $password -AsPlainText -Force
    $credential = New-Object System.Management.Automation.PSCredential ($username, $securePassword)

    Add-ToTrustedHosts -ip $ip

    try {
        $session = New-PSSession -ComputerName ${ip} -Credential $credential -ErrorAction Stop

        Invoke-Command -Session $session -ScriptBlock $remoteScriptBlock

        Remove-PSSession -Session $session
        Log-Info "Restored hosts on ${ip} successfully"
        $successCount++
    }
    catch {
        Log-Error "Failed to restore hosts on ${ip}: $_"
        $failedServers += "$username@${ip}"
        $failureCount++
    }
}

# Calculate failure and success percentages
$totalCount = $servers.Count
if ($totalCount -gt 0) {
    $failurePercentage = [math]::Round(($failureCount / $totalCount) * 100, 2)
    $successPercentage = [math]::Round(($successCount / $totalCount) * 100, 2)
} else {
    $failurePercentage = 0
    $successPercentage = 100
}

# Output summary result and log to file
$summaryContent = @"
========================================
[SUMMARY] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Execution Rollback Summary
========================================
Total number of servers: $totalCount
Number of successful restores: $successCount
Number of failed restores: $failureCount
Failure rate: $failurePercentage%
Success rate: $successPercentage%
----------------------------------------
"@

if ($failedServers.Count -gt 0) {
    $summaryContent += "Failed servers:`n"
    foreach ($server in $failedServers) {
        $summaryContent += "  - $server`n"
    }
}
$summaryContent += "========================================"

# Output summary result to log file and terminal
$summaryContent | Add-Content -Path $summaryLog
Write-Output $summaryContent
Log-Info "Script execution completed. Check $summaryLog for summary."
```


Step 2 Modify the following parameters in the script to meet your needs:

- **$logDir = "C:\Users\Public\Hosts_Script_Logs"**
  - Description: log directory, which is used to store run, error, and summary logs.
  - Default value: **C:\Users\Public\Hosts_Script_Logs**
  - Suggestion: Change the value to a directory for which the current user has the write permission.
  - Example: **$logDir ="C:\Users\username\Documents\Hosts_Script_Logs"**
- **$csvFile = "C:\Users\Public\target_servers.csv"**
  - Description: **CSV file** path. The file contains the source server information.
  - Default value: **C:\Users\Public\target_servers.csv**
  - Suggestion: Use an absolute path or a correct relative path. If the CSV file path changes, you need to update the path here.
  - Example: **$csvFile = "C:\Users\username\Documents\servers.csv"**

Step 3 After the configuration items are modified and saved, run PowerShell as administrator and execute the script.

```
.\rollback_hosts_win.ps1
```

The script will output log information in the terminal window and generate a result report upon completion. You can view this report in the **summary.log** file in the directory specified by **$logDir**.

----End

2.8.4 FAQs

2.8.4.1 How Do I Enable the PowerShell Remoting?

Generally, PowerShell remoting is enabled by default on Windows Server 2012 and later versions. If the settings are changed, you can perform the following steps to enable PowerShell remoting:

Step 1 Run PowerShell as administrator.

Step 2 Run the following command to enable PowerShell remoting:

```
Enable-PSRemoting
```

Step 3 Verify the configuration. Run the following command in PowerShell:

```
New-PSSession
```

- If the configuration is successful, the command creates a session on the local computer and returns a session object. Example output:

```
Id Name            ComputerName    State         ConfigurationName
-- ----            ------------    -----         -----
1  Session1        localhost       Opened        Microsoft.PowerShell
```

- If the configuration fails, refer to **about_Remote_Troubleshooting** in the PowerShell documentation for solutions.

----End

2.8.4.2 How Do I Enable the WinRM Service?

Step 1 Run PowerShell as administrator.

Step 2 Run the following command to automatically start the WinRM service and configure the remote access settings:

```
Enable-PSRemoting -Force
```

----End

2.8.4.3 What Can I If an Error Is Reported After a Script Is Executed, Indicating that the Remote Server Fails to Be Connected and the Login Credential Information Is Correct?

Symptom

After the script was executed, the error message "[ERROR] Failed to update hosts on xxx.xxx.x.xx: [xxx.xxx.x.xx] Failed to connect to the remote server xxx.xxx.x.xx. For details, see the about_Remote_Troubleshooting topic" was displayed.

Possible Causes

The WinRM service on the remote server is not started or is incorrectly configured.

Solution

Step 1 Check whether the executor can ping the remote server. If the ping operation succeeds, go to step 2.

Step 2 Check whether port 5985 used by the WinRM service on the remote server is open to the executor. If it is, go to step 3.

Step 3 Check whether the login credential of the remote server is correct. After confirming that the credential is correct, perform the following steps.

Step 4 Log in to the remote server that fails to be connected and run PowerShell as administrator.

Step 5 Run the following command to run the script again:

```
Enable-PSRemoting -Force
```

----End

3 Reducing Disk Capacity for Target Servers

MgC enables you to reduce disk capacity and quantity for target servers based on the disk usage of source servers. This helps you reduce storage costs.

Precautions

- The system disk capacity ranges from 40 GB to 1,024 GB.
- The data disk capacity ranges from 10 GB to 32,768 GB.
- Disk downsizing is only available for Linux, and the decreased sizes must be larger than the used sizes of the source disks.
- In the cross-AZ migration scenario, only disk upsizing is supported. Even if you choose to downsize disks here, the settings will not be applied, and the system will create target disks as large as source disks.

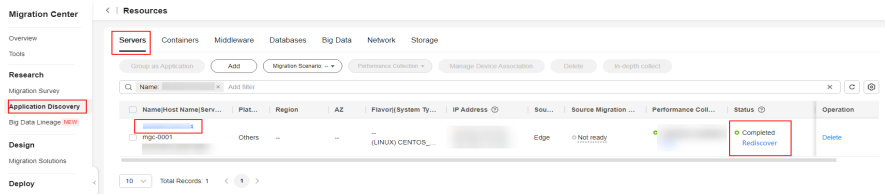
Collecting Disk Information of Source Servers

You need to collect the disk information of source servers and then, against the collected information, reduce disk capacity for target servers.

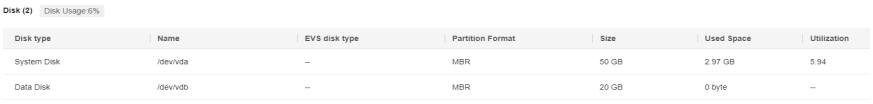
Step 1 Select a collection method based on your requirements.

- **Discovering Resources over the Internet**
- **Discovering Resources over an Intranet**
- **Manually Adding Resources to MgC**

Step 2 Wait for the resource discovery and deep collection to complete. View the server list on the **Resources** page and click a source server.



Step 3 In the disk information area, view the number and usage of disks on the source server. Based on the information, you can adjust disk settings for paired target servers.



----End

Associating Source Servers with Target Servers and Reducing Disk Capacity for Target Servers

Step 1 In the navigation pane, choose **Design** > **Migration Solutions**.

Step 2 Click **View Resources** in the **Target Configuration** card.

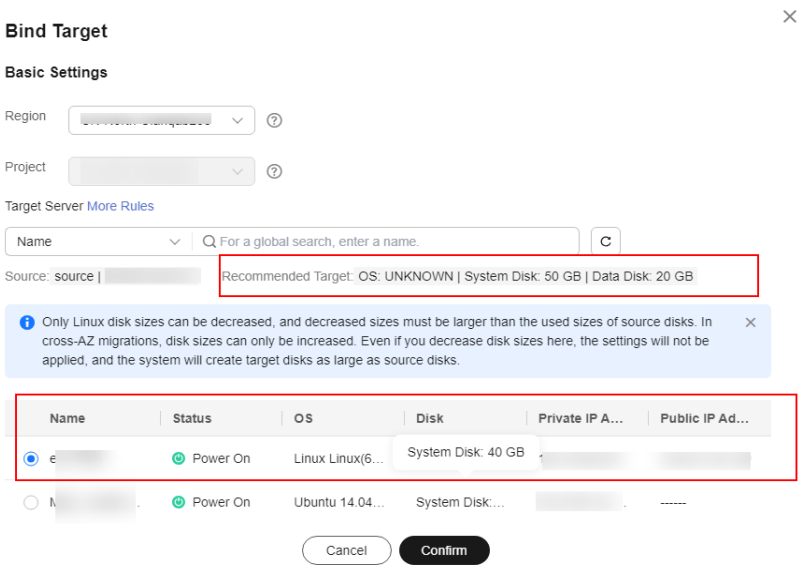
Step 3 On the displayed **Servers** tab, locate a source server and click **Associate** in the **Target Association** column.

Step 4 Select the region of the **application** that the source server was added to, and select a project in that region. In the project, select a target server based on the **collected disk information of the source server** and your requirements.

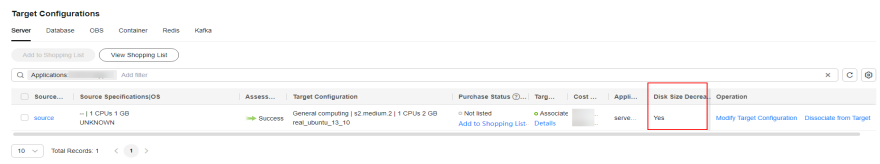
CAUTION

Ensure that the disk capacity of the selected target server is greater than the used disk capacity of the source server.

Assume the source server has a 50 GB system disk with a little space used and a 20 GB data disk that is unused at all. You can associate a target server containing only a 40 GB system disk with the source server.

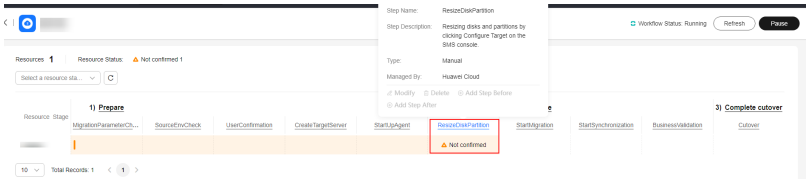


Step 5 Click **Confirm**. The system will automatically check whether the associated target server has downsized disks compared with the source server. If it does, **Yes** will be displayed in the **Disk Downsized** column. If it does not, **No** will be displayed.



Step 6 **Create a server migration workflow**. When the workflow reaches the **ResizeDiskPartition** step, the system identifies whether disk capacity reduction has been performed on the target server.

- If yes, this step is paused. You need to go to **SMS console** and resize disks and partitions for the target server. For details, see **Resizing Disks and Partitions for Target Servers**. After the adjustment is complete, go back to the MgC console and click **Confirm** next to the step status so that the workflow can continue.



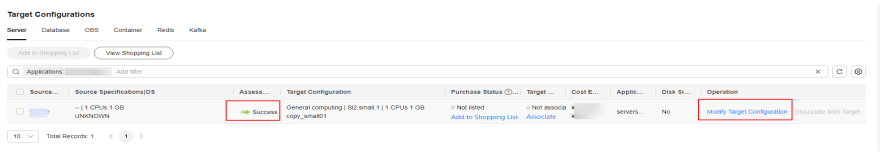
- If no, skip this step and proceed with the subsequent migration steps.

----End

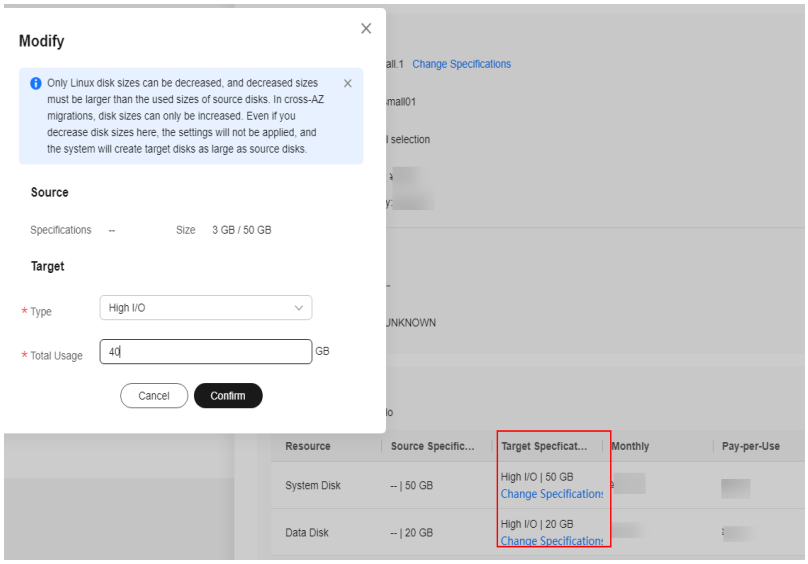
Getting Target Server Recommendations and Reducing Disk Capacity for Target Servers

Step 1 Get recommendations for target servers. For details, see **Getting Target Recommendations**.

Step 2 In the **Target Configurations** area, locate the server that you want to modify the recommended target configurations for and click **Modify Target Configuration** in the **Operation** column.



Step 3 Locate the desired disk and click **Modify** in the **Target Specifications** column.



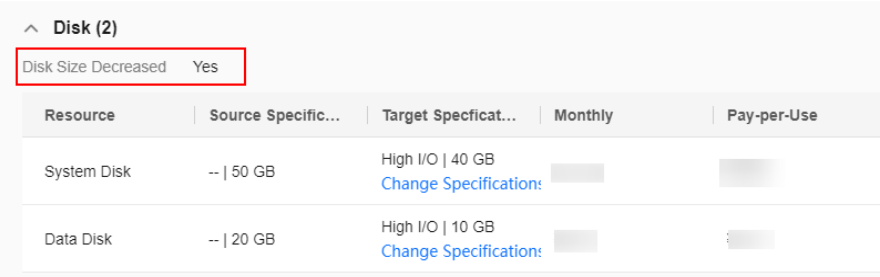
Step 4 Modify the disk capacity based on the **collected disk information of the source server** and your service requirements.

Assume the source server has a 50 GB system disk and a 20 GB data disk, and the usage of both disks is very low. You can reduce the system disk to 40 GB and the data disk to 10 GB for the target server.

CAUTION

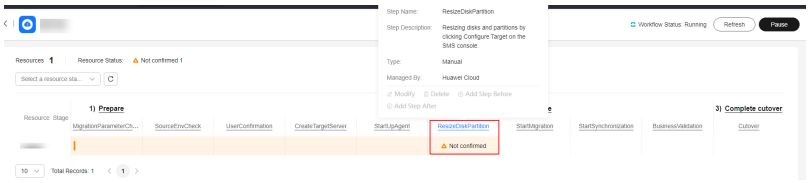
Ensure that the disk capacity of the selected target server is greater than the used disk capacity of the source server.

Step 5 Click **Confirm**. You can see **Yes** is displayed after **Disk Downsized**, which means that the disks are downsized for the target server. If you do not change the disk specifications of the target server, **No** will be displayed after **Disk Downsized**.



Step 6 **Create a server migration workflow**. When the workflow reaches the **ResizeDiskPartition** step, the system identifies whether disk capacity reduction has been performed on the target server.

- If yes, this step is paused. You need to go to **SMS console** and resize disks and partitions for the target server. For details, see **Resizing Disks and Partitions for Target Servers**. After the adjustment is complete, go back to the MgC console and click **Confirm** next to the step status so that the workflow can continue.



- If no, skip this step and proceed with the subsequent migration steps.

----End

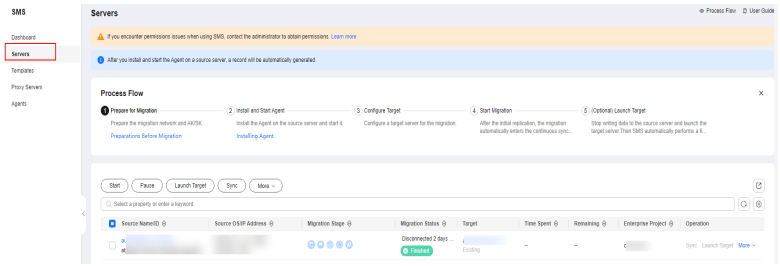
4 Resizing Disks and Partitions for Target Servers

In a server migration workflow, if the system detects that the disk capacity reduction has been performed on a target server, the workflow will be paused, and you need to go to the SMS console to resize disks and partitions for the target server.

Procedure

Step 1 Sign in to the **SMS console**.

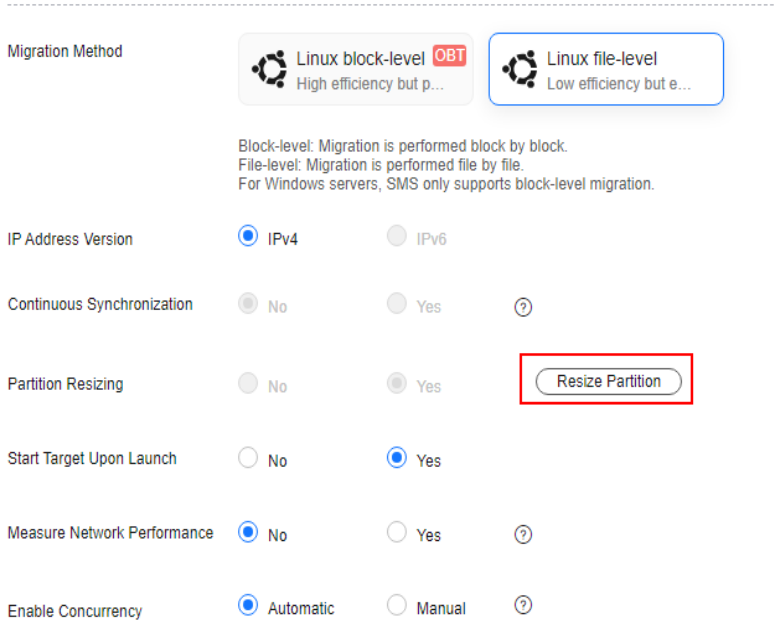
Step 2 In the navigation pane on the left, choose **Servers**.



Step 3 Locate the desired server in the server list based on the resource name in the MgC migration workflow, and click **Configure** in the **Target** column.

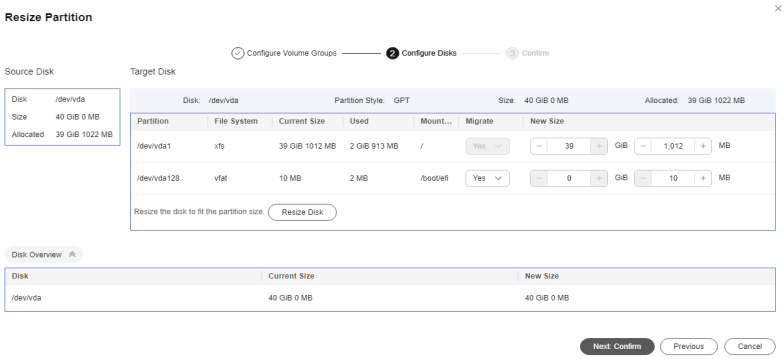
Step 4 Select **Configure now** next to **Advanced Settings**.

Step 5 Click **Resize Partition** next to **Partition Resizing**.



Step 6 Adjust the disk size, disk quantity, and partition size based on the target server specifications configured in the workflow.

Figure 4-1 Resizing disks and partitions on Linux



NOTE

- For a Linux server using LVM, you can choose whether to migrate physical or logical volumes and resize the paired target volumes.
- Partition resizing is not available for Btrfs partitions on Linux.
- In a Linux migration, the system and swap partitions are migrated by default.
- You can choose to migrate all or none volume groups by configuring **Migrate All Volume Groups**.
- If you choose to migrate none of the logical volumes in a volume group, their physical volumes will not be migrated by default.

Step 7 After the configuration is completed, click **Next: Confirm**. After confirming that the configuration is correct, click **OK**.

Step 8 Click **Next: Configure Target** in the lower right corner.

Step 9 In the server list, select the target server paired with the source server and click **Next: Confirm**. You can view the name of the target server by clicking the **CreateTargetServer** step in the MgC migration workflow.

Step 10 After confirming that the configuration is correct, click **Save**. Read the migration checklist carefully and click **OK**.

Step 11 Return to the MgC migration workflow. Locate the **ResizeDiskPartition** step, and click **Confirm** next to the step status to continue the subsequent migration steps.

----End

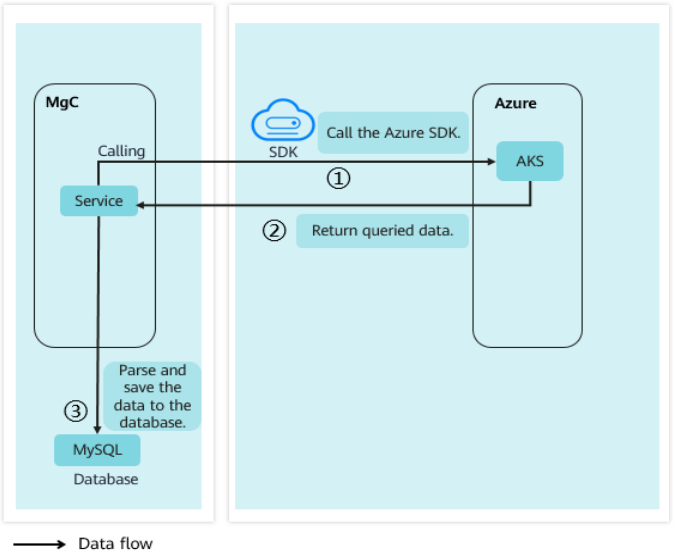
5 Collecting Details of Azure Kubernetes Service (AKS) Resources

Before migrating Azure Kubernetes Service (AKS) resources, use MgC to collect resource details, which are necessary for subsequent migration. This section describes the basic principles, preparations, account permission requirements, and specific operations for using MgC to collect AKS resource details efficiently and accurately.

Collection Principles

Figure 5-1 illustrates how to use MgC to collect AKS resource details.

Figure 5-1 Principle for collecting AKS resource details



1. MgC invokes AKS APIs through the Azure SDK to obtain information about resources such as containers and VMs.
2. MgC receives API call responses, which typically contain extensive resource data.

3. MgC parses the returned data and extracts key information, such as the number of nodes and number of VM cores. Then, the key information is saved to the database for subsequent analysis and migration.

Preparations

- Preparing a Huawei account

  Before using MgC, you need to prepare a HUAWEI ID or an IAM user that can access MgC. For details about how to register a HUAWEI ID and create an IAM user, see **Preparations**.

- Creating an application migration project

  Create a migration project (a simple project is recommended) on the MgC console. For details, see **Managing Migration Projects**.

- Preparing Azure credentials

  Obtain the password of the application client that owns the AKS resources, subscription ID used to purchase the AKS resources, tenant ID of the application, and client (application) ID. To learn how to obtain Azure credentials, see **How Do I Obtain Azure Credentials?**

- Providing the source credentials

  Add the Azure authentication information to the MgC console as the collection credential. For details, see **Managing Credentials**.

Required Permissions

Ensure that the application to which the added Azure credentials belong has the following permissions in the resource group and subscription for purchasing the AKS resources:

- Microsoft.ClassicCompute/virtualMachines/read
- Microsoft.Insights/MetricDefinitions/Read
- Microsoft.Management/getEntities/action

For details, see **How Do I Configure the Permissions Required for Collecting Details of Azure Containers?**

Procedure

Create a discovery task on the MgC console. For details, see **Discovering Resources over the Internet**.

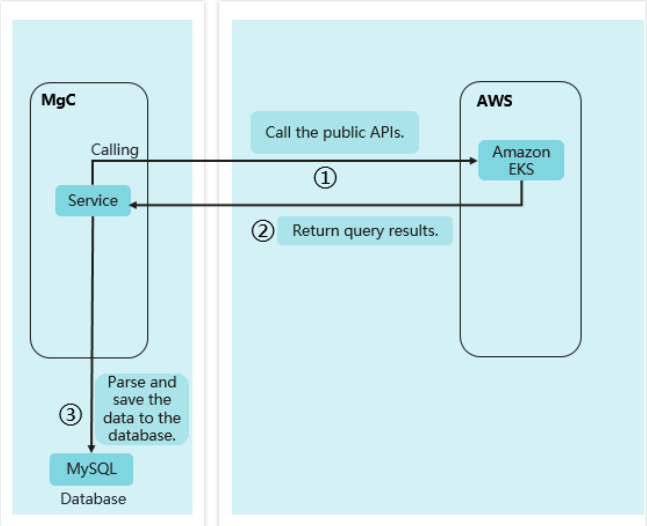
6 Collecting Details of AWS Container Resources

This section describes the basic principles, preparations, account permission requirements, and specific operations for using MgC to collect AWS container resource details efficiently and accurately.

Principle of Collection over the Internet

Figure 6-1 illustrates how to use MgC to collect details about AWS container resources over the Internet.

Figure 6-1 Principle of collecting AWS container resource details

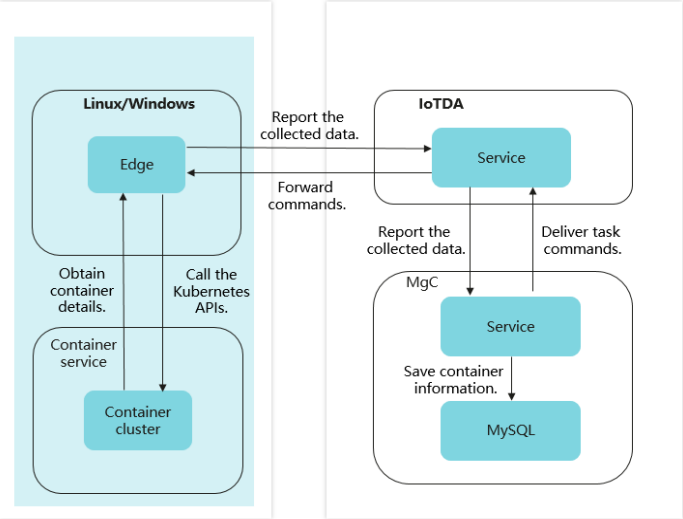


1. MgC invokes Amazon EKS APIs to obtain information about resources such as containers and VMs.
2. MgC receives API call responses, which typically contain extensive resource data.
3. MgC parses the returned data and extracts key information, such as the number of nodes and number of VM cores. Then, the key information is saved to the database for subsequent analysis and migration.

Principle of Deep Collection

Figure 6-2 shows the principle of a deep collection for AWS container resources by MgC.

Figure 6-2 Principle of deep collection for AWS containers



The process is as follows:

1. MgC sends commands to the MgC Agent to collect container resource information.
2. The MgC Agent accesses the container cluster using the credentials you provide.
3. The MgC Agent calls Kubernetes APIs to collect cluster details, including container specifications, node configurations, persistent volume configurations, and network policies.
4. The MgC Agent reports the collected information to MgC.
5. After receiving the reported information, MgC parses the information, extracts useful information, and saves the information to the database.

Preparations

- Preparing a Huawei account

  Before using MgC, you need to prepare a HUAWEI ID or an IAM user that can access MgC. For details about how to register a HUAWEI ID and create an IAM user, see **Preparations**.

- Creating an application migration project

  Create a migration project (a simple project is recommended) on the MgC console. For details, see **Managing Migration Projects**.

- Preparing AWS account credentials

  Obtain an AK/SK pair for the AWS account that owns the resources to be collected. For details, see **Obtaining AWS Access Keys**.

- Providing the source credentials

Add the AWS authentication information to the MgC console as the collection credential. For details, see **Managing Credentials**.

- Obtaining the login configuration files of the AWS container clusters

  The configuration files are used for deep collection. The MgC Agent uses them to access the AWS container clusters and invoke the Kubernetes APIs to collect details about the container clusters.

Required Permissions

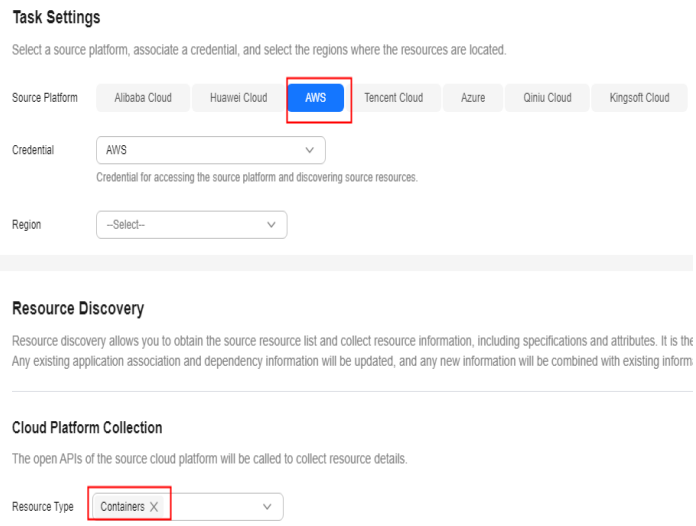
Before collecting details of Amazon EKS resources, ensure that the AWS account has the following permissions:

- eks:DescribeCluster
- eks:ListClusters
- ec2:DescribeInstances
- ec2:DescribeSubnets
- cloudwatch:GetMetricStatistics

Creating an Internet-based Discovery Task

Create a discovery task on the MgC console. For details, see **Discovering Resources over the Internet**. During the task creation, set **Source Platform** to **AWS**, **Credential** to the credential provided in preparations, **Region** as required, and **Resource Type** to **Container**.

Figure 6-3 Creating an Internet-based discovery task

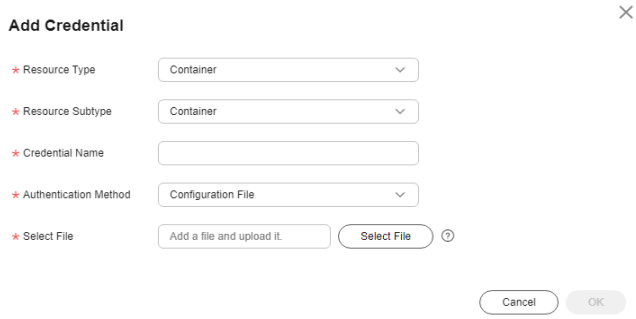


Performing a Deep Collection

After obtaining the list of AWS containers through the **Internet-based discovery task**, you can perform a deep collection for container resources to obtain their details, including container specifications, node configurations, persistent volume configurations, and network policies. For details, see **Performing a Deep Collection for Containers**.

Before that, you need to provide the MgC Agent with the login configuration files for accessing the AWS container clusters.

Figure 6-4 Adding credentials required for deep collection to the MgC Agent



7 Verifying Big Data Consistency After Migration

7.1 Verifying the Consistency of Data Migrated from MaxCompute to DLI

This section describes how to use MgC to verify the consistency of data migrated from Alibaba Cloud MaxCompute to Huawei Cloud Data Lake Insight (DLI).

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in to the tool. For details, see **Installing the MgC Agent on Linux**.

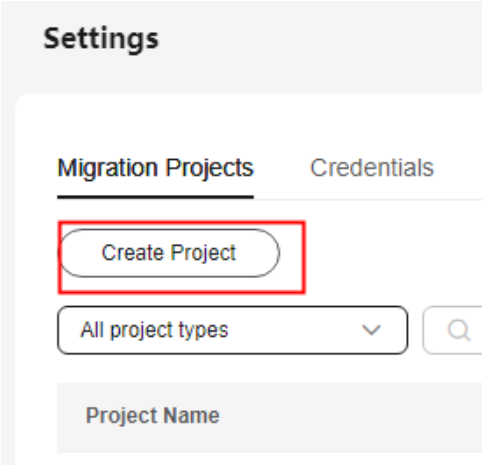
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

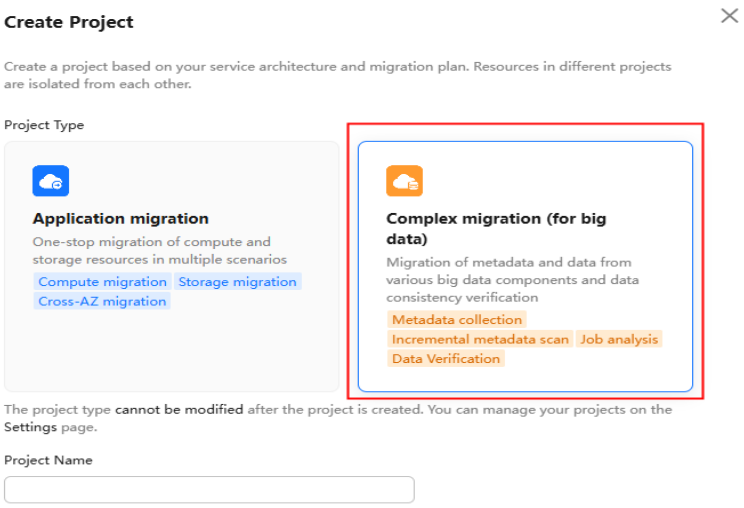
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-1 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-2 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 On the MgC Agent console, add your AK/SK pairs required for accessing MaxCompute and DLI. For more information, see **Adding Resource Credentials**.

- For details about how to obtain an AK/SK pair for accessing DLI, see **How Do I Obtain the AK/SK Pair?**
- For details about how to obtain an AK/SK pair for accessing MaxCompute, see **Viewing the Information About AccessKey Pairs of a RAM User**.

Step 7 In the navigation pane, choose **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

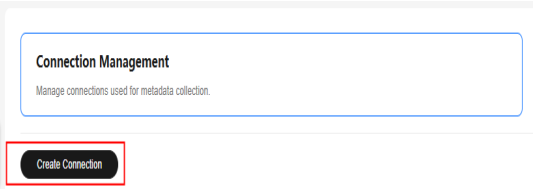
Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

> **CAUTION**
>
> Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

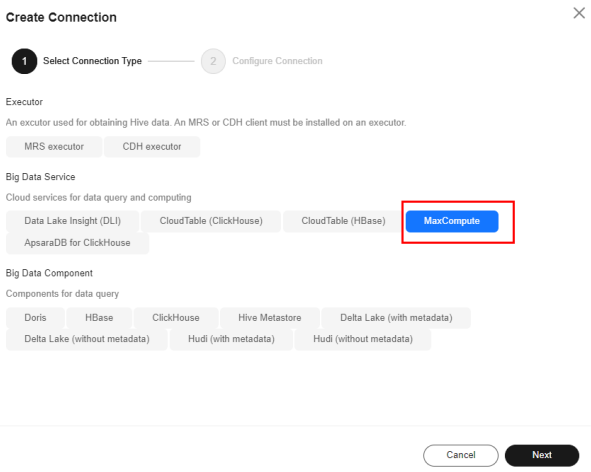
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-3 Creating a connection



Step 11 On the **Select Connection Type** page, select **MaxCompute** and click **Next**.



Step 12 Configure the **parameters for creating a connection to MaxCompute**, and click **Test**. If the test is successful, the connection is set up.

Table 7-1 Parameters for creating a connection to MaxCompute

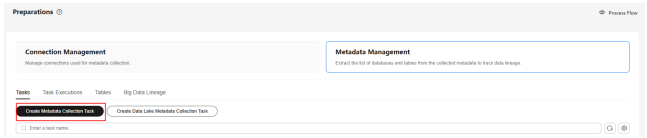
Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **MaxCompute-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Alibaba Cloud Credential	Select the MaxCompute credential added to the MgC Agent in **step 6**.

Parameter	Configuration
MaxCompute Project	Enter the name of your MaxCompute project. You can obtain the project name from the MaxCompute console.
Endpoint	Enter the endpoint of the region where the MaxCompute project is located. For details about the MaxCompute endpoints in different regions, see **MaxCompute Endpoints**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-4 Create Metadata Collection Task



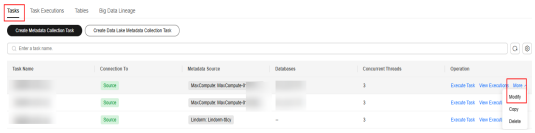
Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

Table 7-2 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.
Databases	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. **NOTICE** This parameter is mandatory only if a MaxCompute metadata connection is selected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent (formerly Edge) resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

Figure 7-5 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-6 Managing task executions

Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

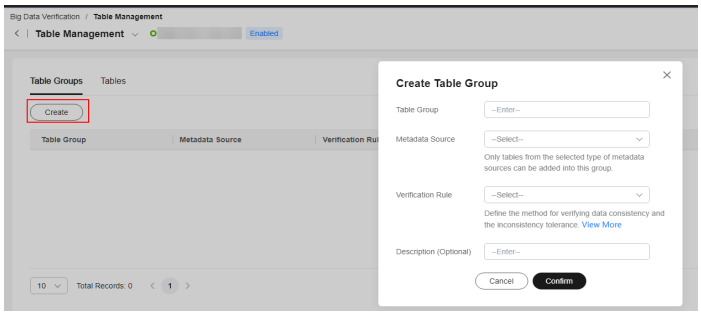
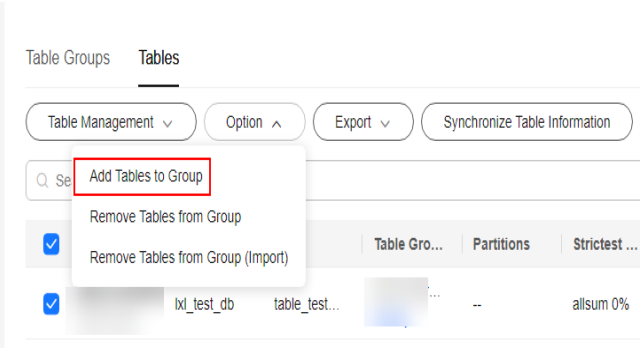


Table 7-3 Parameters for creating a table group

Parameter	Configuration
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see more information about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



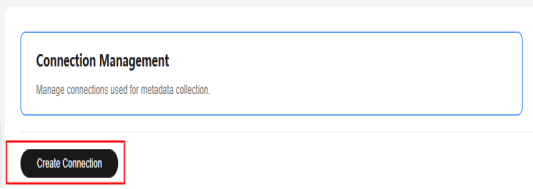
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

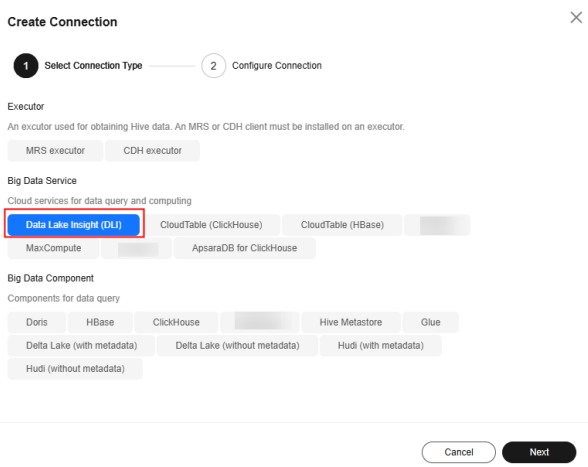
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-7 Creating a connection



Step 24 On the **Select Connection Type** page, select **Data Lake Insight (DLI)** and click **Next**.



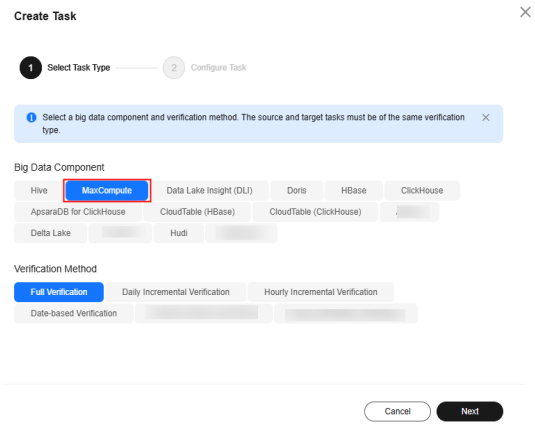
Step 25 Configure the **parameters required for creating a connection to DLI**, and click **Test**. If the test is successful, the connection is set up.

Table 7-4 Parameters for creating a DLI connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **DLI-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
DLI Credential	Select the DL credential added to the MgC Agent in **step 6**. If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed.
Region ID	Enter the code of the target region where the data to be verified is located, for example, ap-southeast-1. For details about region codes, see **Endpoints**.
Project ID	Enter the ID of the project where the data to be verified is stored. For details about how to obtain the project ID, see **Obtaining Project Information**.
Queue	Enter the name of the DLI queue used to execute verification. The queue must be a SQL queue.
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI.

Step 26 On the MgC console, create a verification task for MaxCompute and execute the task. For details, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

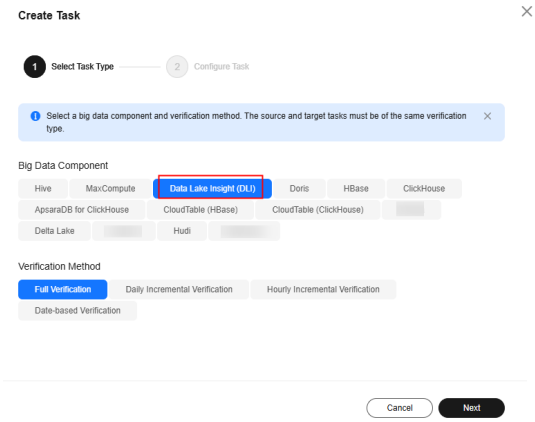
- On the **Select Task Type** page, select **MaxCompute** for **Big Data Component**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 27 On the MgC console, create a verification task for DLI and execute the task. For details, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **Data Lake Insight (DLI)**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 28 Wait until the task executions enter a **Completed** status. On the **Verification Results** page, you can view and export the task execution results. For details, see **Viewing and Exporting Verification Results**.

----End

7.2 Verifying the Consistency of Data Migrated Between MRS ClickHouse Clusters

This section describes how to use MgC to verify the consistency of data migrated between MRS ClickHouse clusters of different versions.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

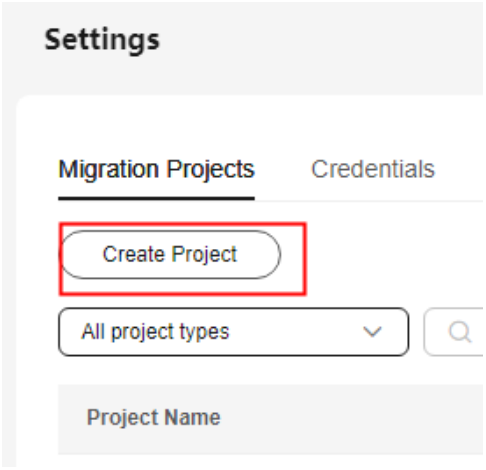
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

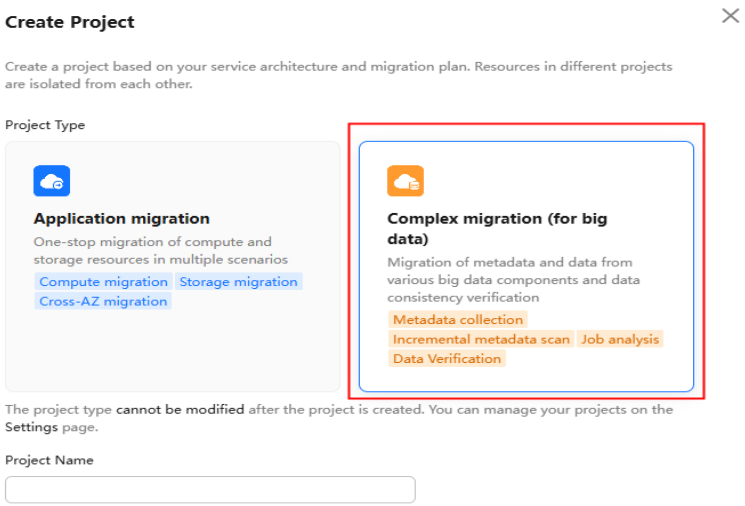
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-8 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-9 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source and target MRS ClickHouse clusters to the MgC Agent. For more information, see **Adding Resource Credentials**.

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

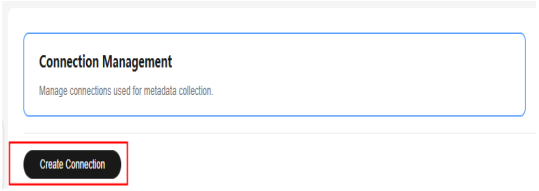
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

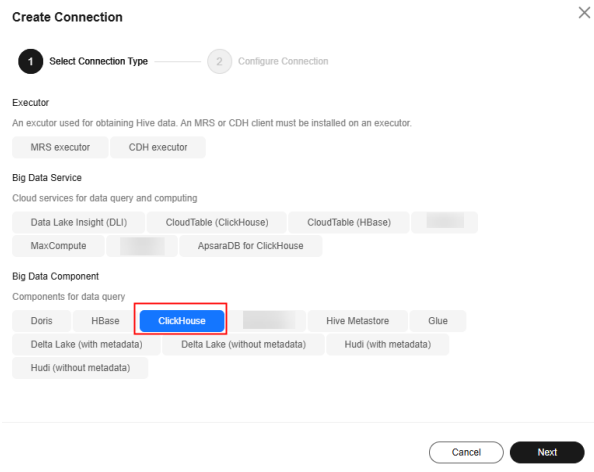
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-10 Creating a connection



Step 11 On the **Select Connection Type** page, select **ClickHouse** and click **Next**.



Step 12 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

Table 7-5 Parameters for creating a connection to ClickHouse

Parameter	Configuration
Connection To	Select **Source**.

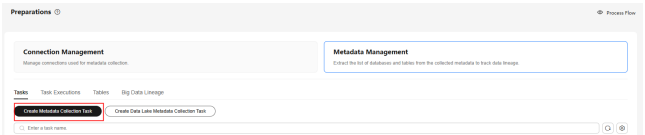
Parameter	Configuration
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential you added to the MgC Agent for accessing the source MRS ClickHouse cluster in **step 6**.
Secured Cluster	Choose whether the cluster is secured.
ClickHouse Server IP Address	Enter the IP address for accessing the source ClickHouse server. Generally, the IP address refers to that of the server where ClickHouse is hosted.
HTTP Port	If the ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the source cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **http_port** parameter.
HTTP SSL/TLS Port	If the ClickHouse cluster is secured, enter the HTTPS port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager**, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **https_port** parameter.

Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-11 Create Metadata Collection Task



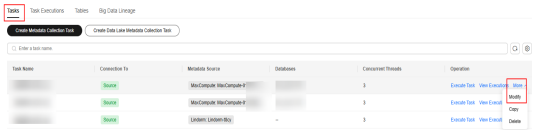
Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

Table 7-6 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

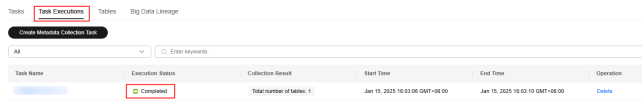
Figure 7-12 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-13 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

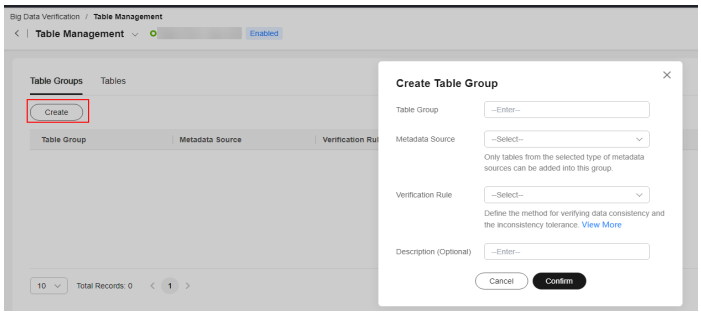
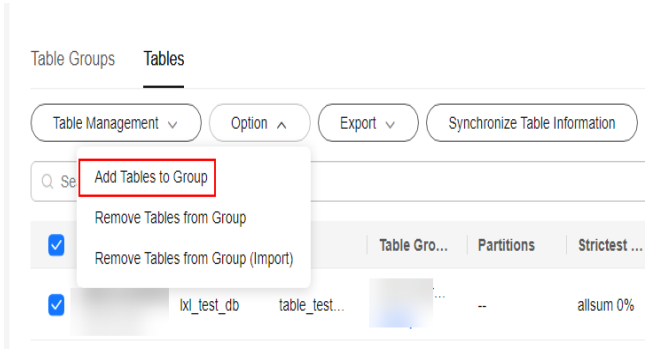


Table 7-7 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see more information about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



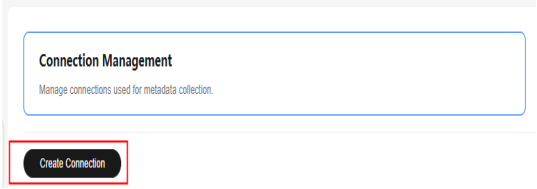
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

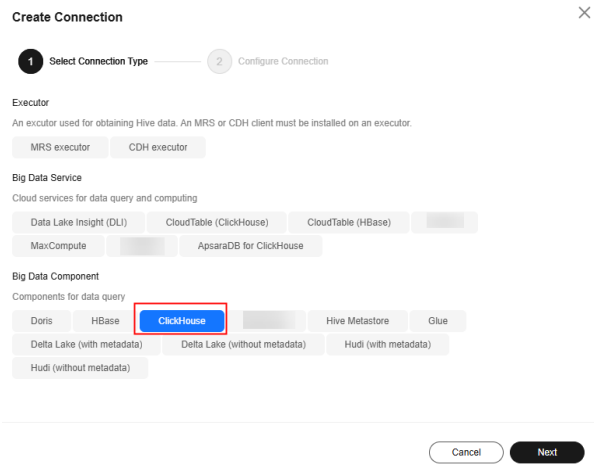
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-14 Creating a connection



Step 24 On the **Select Connection Type** page, select **ClickHouse** and click **Next**.



Step 25 Configure the **parameters for creating a ClickHouse connection**, and click **Test**. If the test is successful, the connection is set up.

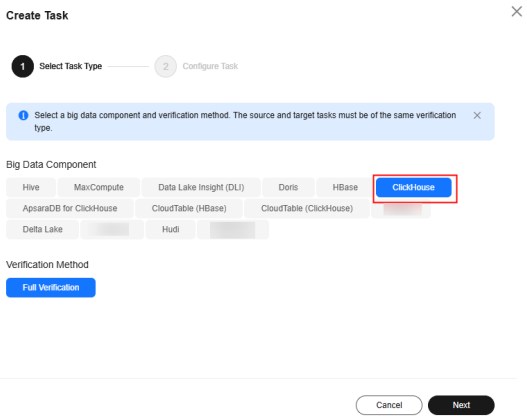
Table 7-8 Parameters for creating a ClickHouse connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential you added to the MgC Agent for accessing the target MRS ClickHouse cluster in **step 6**.
Secured Cluster	Choose whether the cluster is secured.
ClickHouse Server IP Address	Enter the IP address of the MRS ClickHouse server. Generally, the IP address refers to that of the server where ClickHouse is hosted.

Parameter	Configuration
HTTP Port	If the MRS ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the target cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **http_port** parameter.
HTTP SSL/TLS Port	If the MRS ClickHouse cluster is secured, enter the HTTPS port for communicating with the MRS ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the target cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **https_port** parameter.
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 26 Create a data verification task for the source and target MRS ClickHouse clusters, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, select **ClickHouse** for **Big Data Component**.



- Select **Full Verification** for **Verification Method**.

Step 27 Wait until the task executions enter a **Completed** status. On the **Verification Results** page, you can view and export the task execution results. For details, see **Viewing and Exporting Verification Results**.

----End

7.3 Verifying the Consistency of Data Migrated from Alibaba Cloud EMR ClickHouse to Huawei Cloud MRS ClickHouse

This section describes how to use MgC to verify the consistency of data migrated from Alibaba Cloud EMR ClickHouse to Huawei Cloud MRS ClickHouse.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

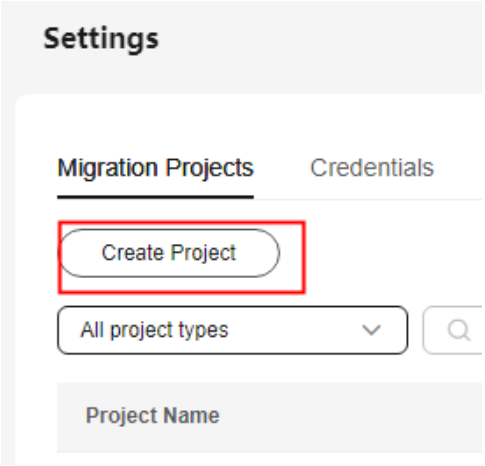
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

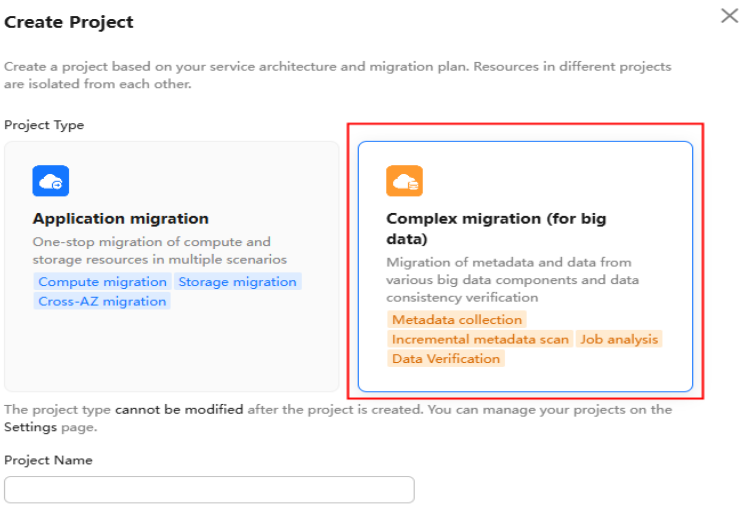
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-15 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-16 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source and target ClickHouse servers to the MgC Agent. For more information, see **Adding Resource Credentials**.

To obtain the username and password for logging in to the Alibaba Cloud EMR ClickHouse server, go to the EMR console, on the **Configuration** page of ClickHouse, and view the **users.default.password** parameter.

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

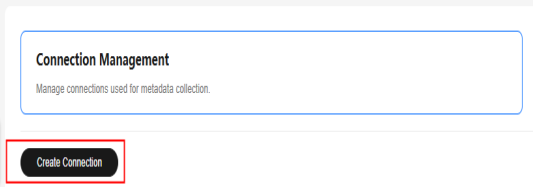
Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

> **CAUTION**
>
> Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

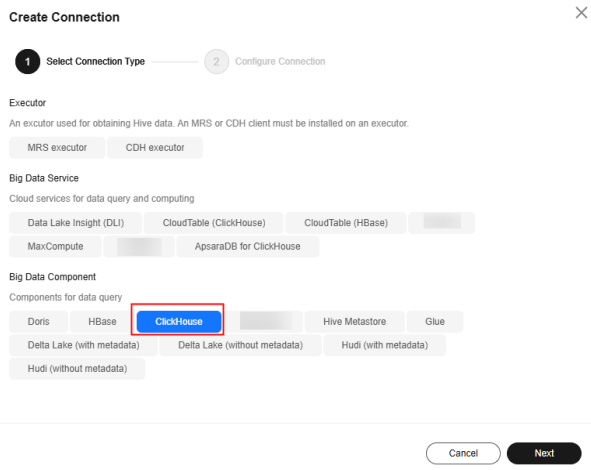
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-17 Creating a connection



Step 11 On the **Select Connection Type** page, select **ClickHouse** and click **Next**.



Step 12 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

Table 7-9 Parameters for creating a connection to ClickHouse

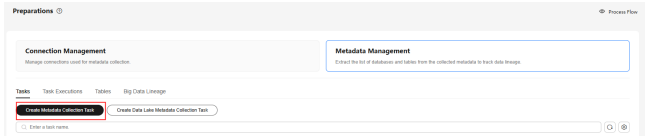
Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential you added to the MgC Agent for accessing the Alibaba Cloud MRS ClickHouse server in **step 6**.
Secured Cluster	Choose whether the cluster is secured.

Parameter	Configuration
ClickHouse Server IP Address	Enter the IP address of the Alibaba Cloud EMR ClickHouse server. Generally, the IP address refers to that of the server where ClickHouse is hosted.
HTTP Port	If the Alibaba Cloud MRS ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. To obtain the value, log in to the EMR console, go to the **Configuration** page of the ClickHouse service, click the **server-config** tab, and view the value of **http_port**.
HTTP SSL/TLS Port	If the Alibaba Cloud MRS ClickHouse cluster is unsecured, enter the HTTPS port for communicating with the ClickHouse server. To obtain the value, log in to the EMR console, go to the **Configuration** page of the ClickHouse service, click the **server-config** tab, and view the value of **http_port**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-18 Create Metadata Collection Task



Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

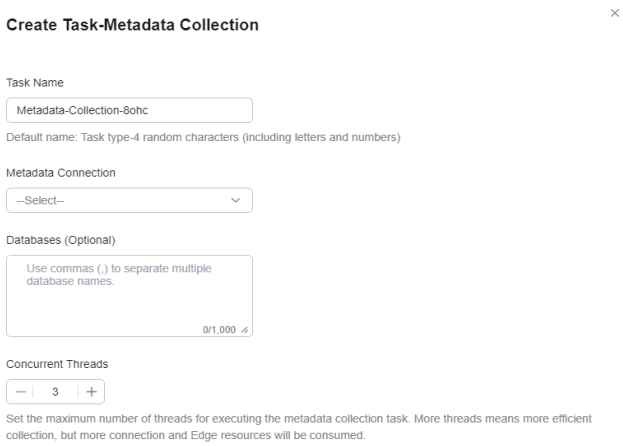
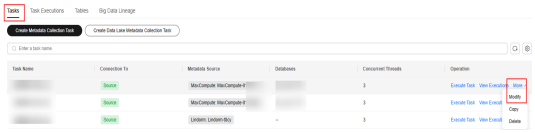


Table 7-10 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

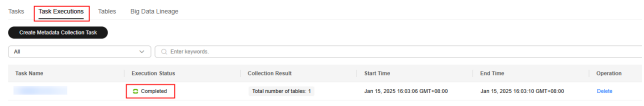
Figure 7-19 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-20 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

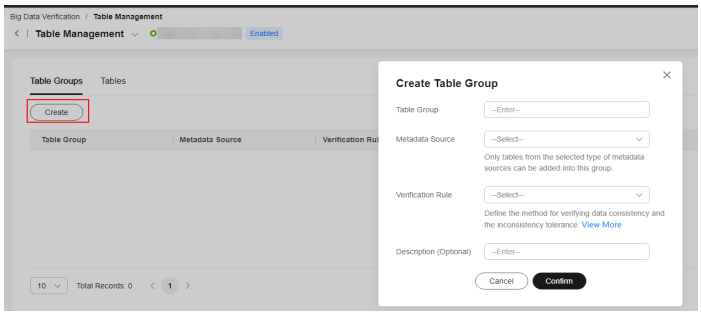
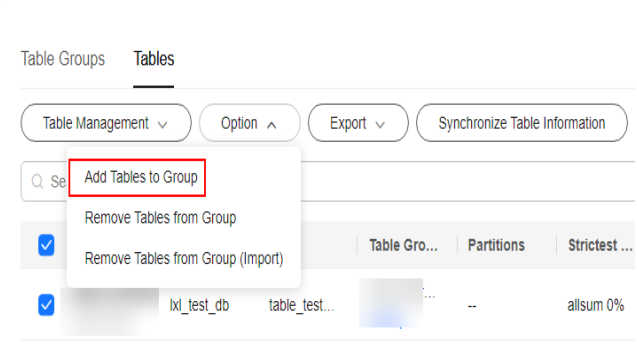


Table 7-11 Parameters for creating a table group

Parameter	Description
Table Group	User-defined
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see more information about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



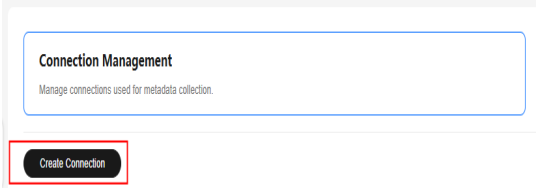
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

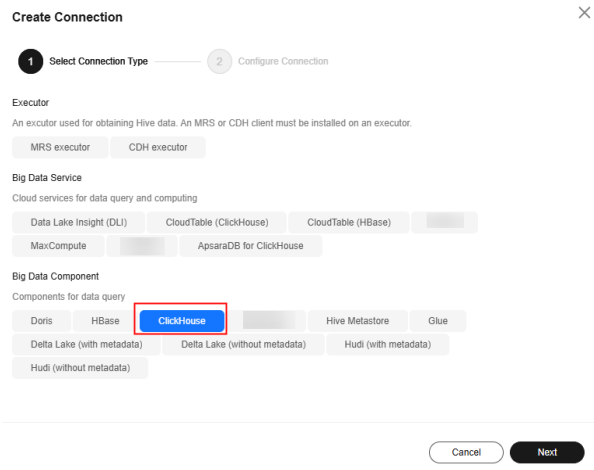
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-21 Creating a connection



Step 24 On the **Select Connection Type** page, select **ClickHouse** and click **Next**.



Step 25 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

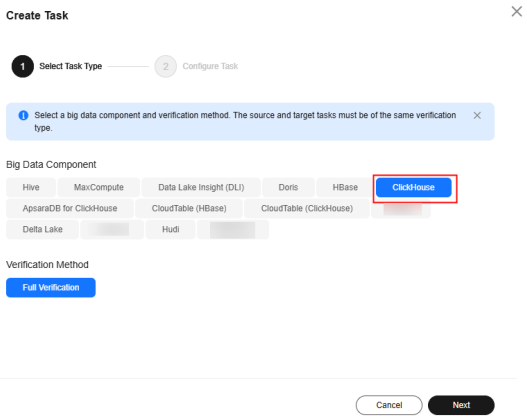
Table 7-12 Parameters for creating a ClickHouse connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential you added to the MgC Agent for accessing the target MRS ClickHouse cluster in **step 6**.
Secured Cluster	Choose whether the cluster is secured.
ClickHouse Server IP Address	Enter the IP address of the MRS ClickHouse server. Generally, the IP address refers to that of the server where ClickHouse is hosted.

Parameter	Configuration
HTTP Port	If the MRS ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the target cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **http_port** parameter.
HTTP SSL/TLS Port	If the MRS ClickHouse cluster is secured, enter the HTTPS port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the target cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **https_port** parameter.
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 26 Create a data verification task for the source EMR ClickHouse cluster and the target MRS ClickHouse cluster, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, select **ClickHouse** for **Big Data Component**.



- Select **Full Verification** for **Verification Method**.

Step 27 Wait until the task executions enter a **Completed** status. On the **Verification Results** page, you can view and export the task execution results. For details, see **Viewing and Exporting Verification Results**.

----End

7.4 Verifying the Consistency of Data Migrated from Alibaba Cloud ApsaraDB for ClickHouse to Huawei Cloud MRS ClickHouse

This section describes how to use MgC to verify the consistency of data migrated from Alibaba Cloud ApsaraDB for ClickHouse to Huawei Cloud MRS ClickHouse.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

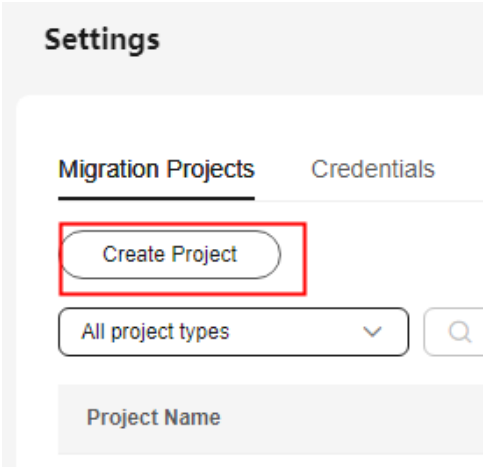
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

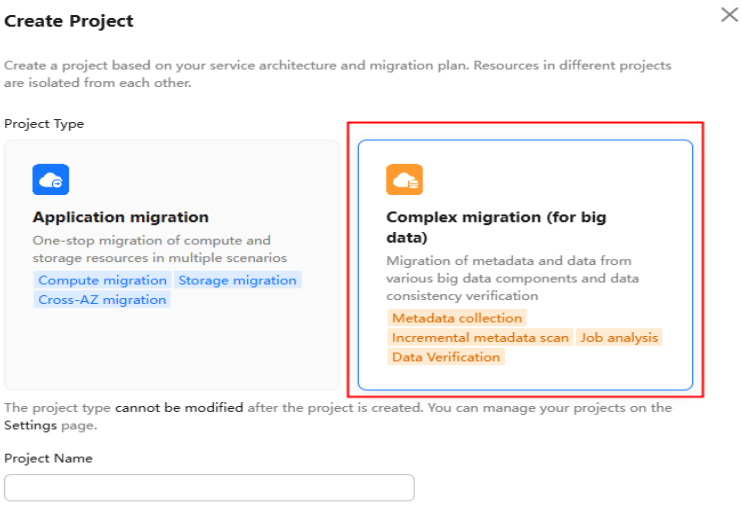
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-22 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-23 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source and target ClickHouse servers to the MgC Agent. For more information, see **Adding Resource Credentials**.

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

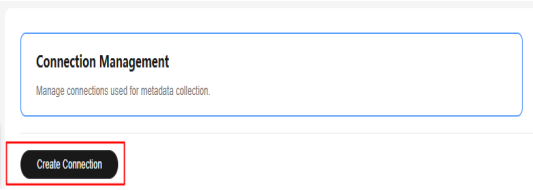
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

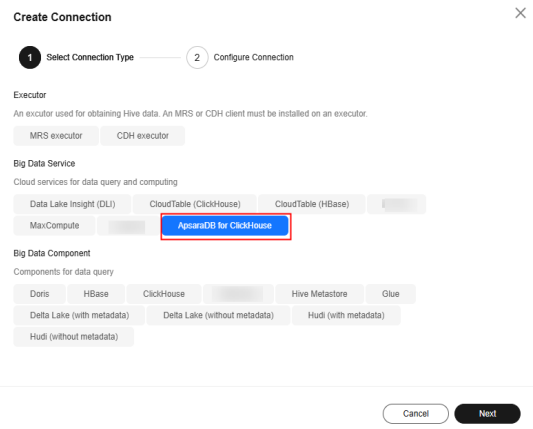
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-24 Creating a connection



Step 11 On the **Select Connection Type** page, select **ApsaraDB for ClickHouse** and click **Next**.



Step 12 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

Table 7-13 Parameters for creating a connection to ClickHouse

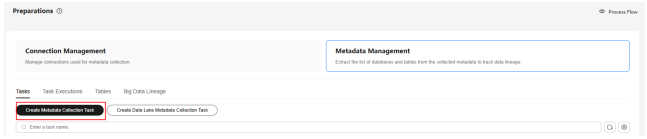
Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **ApsaraDB for ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential used for accessing Alibaba Cloud ApsaraDB for ClickHouse you added to the MgC Agent in **step 6**.

Parameter	Configuration
Database URL	Enter the public address of the source ClickHouse cluster. You can view the IP address in the cluster details.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-25 Create Metadata Collection Task



Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

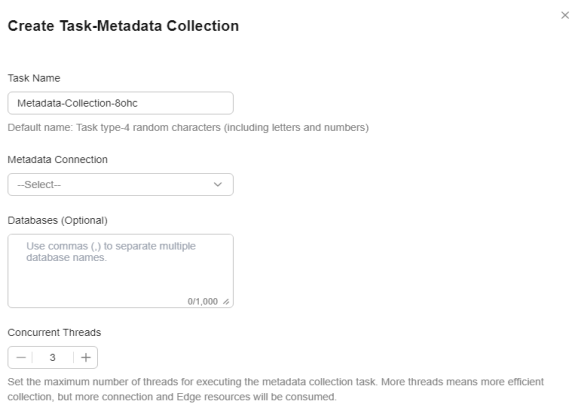


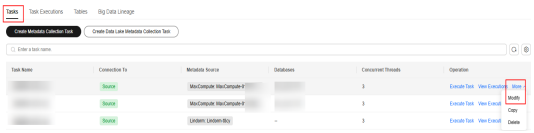
Table 7-14 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. If no database name is specified, the metadata of all databases is collected.

Parameter	Configuration
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

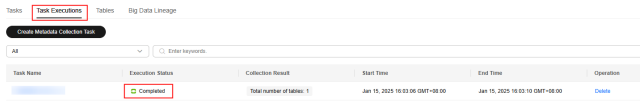
Figure 7-26 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-27 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

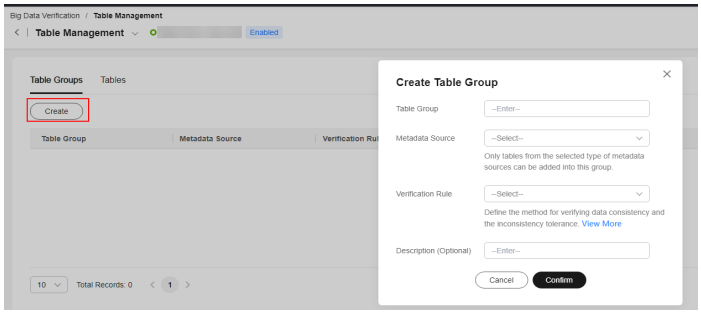
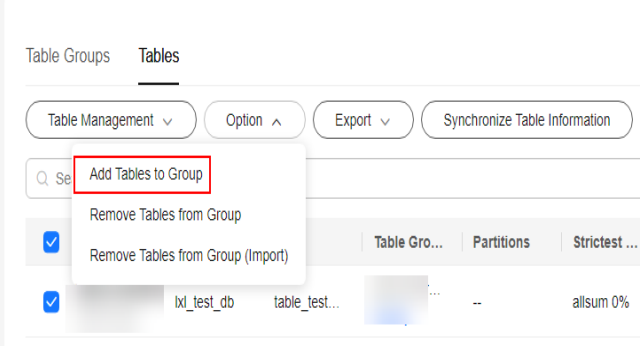


Table 7-15 Parameters for creating a table group

Parameter	Description
Table Group	User-defined

Parameter	Description
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see more information about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



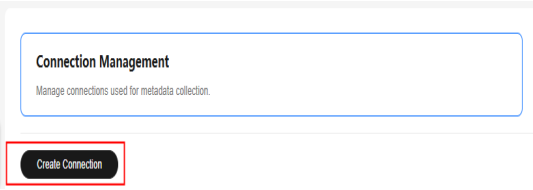
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

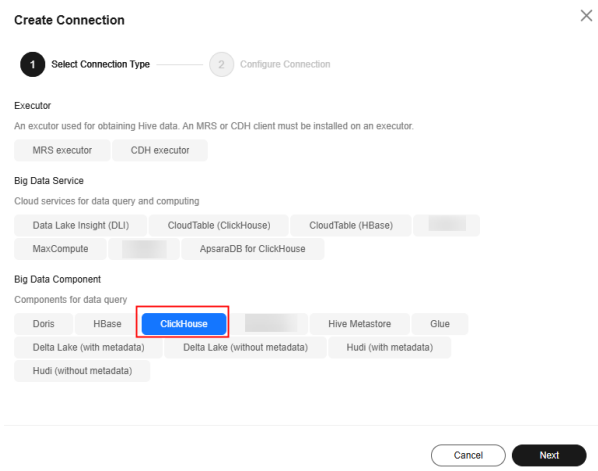
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-28 Creating a connection



Step 24 On the **Select Connection Type** page, select **ClickHouse** and click **Next**.



Step 25 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

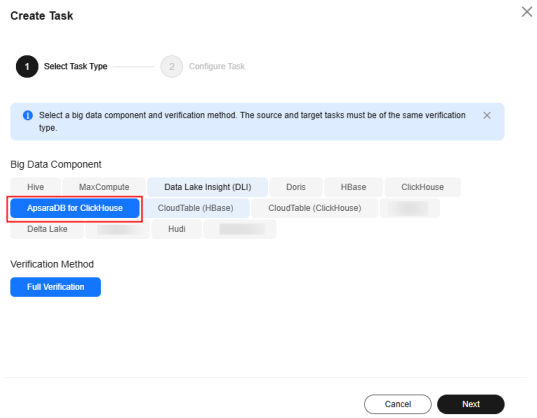
Table 7-16 Parameters for creating a ClickHouse connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential you added to the MgC Agent for accessing the target MRS ClickHouse cluster in **step 6**.
Secured Cluster	Choose whether the cluster is secured.
ClickHouse Server IP Address	Enter the IP address of the MRS ClickHouse server. Generally, the IP address refers to that of the server where ClickHouse is hosted.
HTTP Port	If the MRS ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the target cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **http_port** parameter.
HTTP SSL/TLS Port	If the MRS ClickHouse cluster is secured, enter the HTTPS port for communicating with the ClickHouse server. To obtain the value, log in to the **FusionInsight Manager** of the target cluster, choose **Cluster** > **Services** > **ClickHouse** > **Configurations** > **All Configurations**, and search for the **https_port** parameter.

Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 26 Create a data verification task for the source Alibaba Cloud ApsaraDB for ClickHouse cluster, and execute the task. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

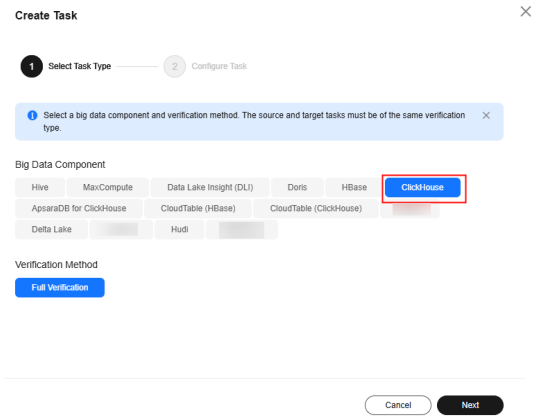
- On the **Select Task Type** page, choose **ApsaraDB for ClickHouse**.



- Select **Full Verification** for **Verification Method**.

Step 27 Create a data verification task for the MRS ClickHouse cluster, and execute the task. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, select **ClickHouse** for **Big Data Component**.



- Select **Full Verification** for **Verification Method**.

Step 28 Wait until the task executions enter a **Completed** status. On the **Verification Results** page, you can view and export the task execution results. For details, see **Viewing and Exporting Verification Results**.

----End

7.5 Verifying the Consistency of Data Migrated from Alibaba Cloud ApsaraDB for ClickHouse to Huawei Cloud CloudTable (ClickHouse)

This section describes how to use MgC to verify the consistency of data migrated from Alibaba Cloud ApsaraDB for ClickHouse to Huawei Cloud CloudTable (ClickHouse).

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

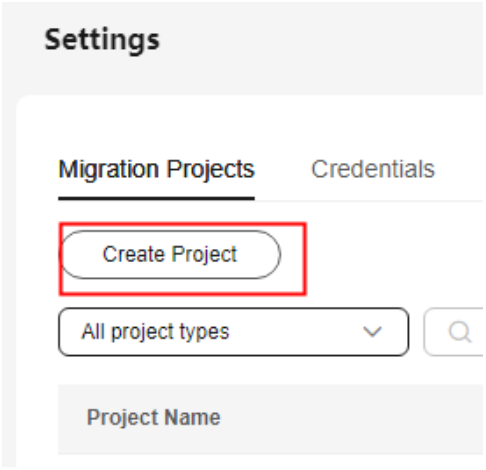
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

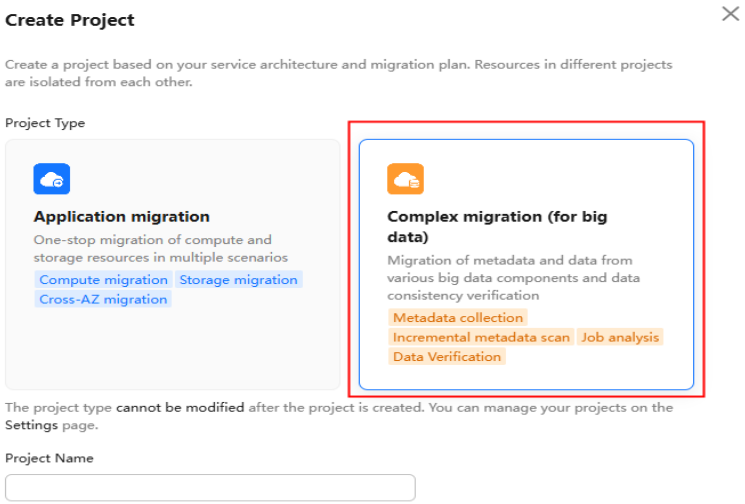
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-29 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-30 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source and target ClickHouse servers to the MgC Agent. For more information, see **Adding Resource Credentials**.

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

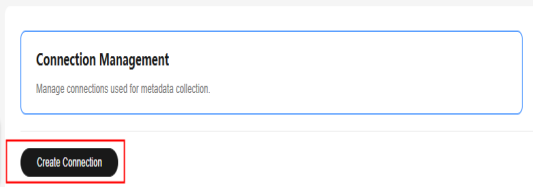
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

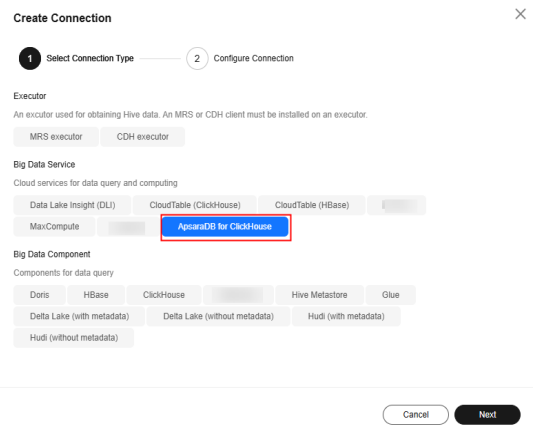
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-31 Creating a connection



Step 11 On the **Select Connection Type** page, select **ApsaraDB for ClickHouse** and click **Next**.



Step 12 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

Table 7-17 Parameters for creating a connection to ClickHouse

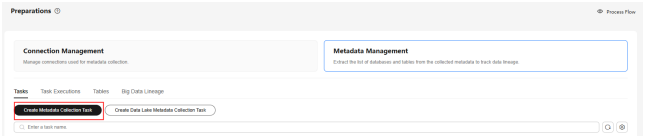
Parameter	Configuration
Connection To	Select **Source**.

Parameter	Configuration
Connection Name	The default name is **ApsaraDB for ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
ClickHouse Credential (Optional)	Select the credential you added to the MgC Agent for accessing the source ApsaraDB for ClickHouse cluster in **step 6**.
Database URL	Enter the public address of the source ClickHouse cluster. You can view the IP address in the cluster details.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-32 Create Metadata Collection Task



Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

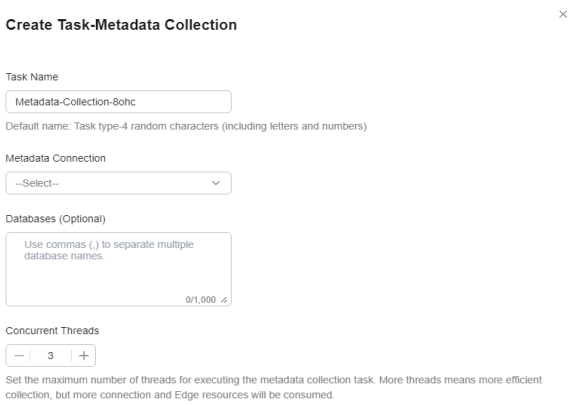


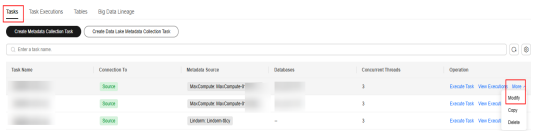
Table 7-18 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.

Parameter	Configuration
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

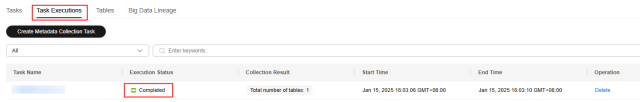
Figure 7-33 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-34 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

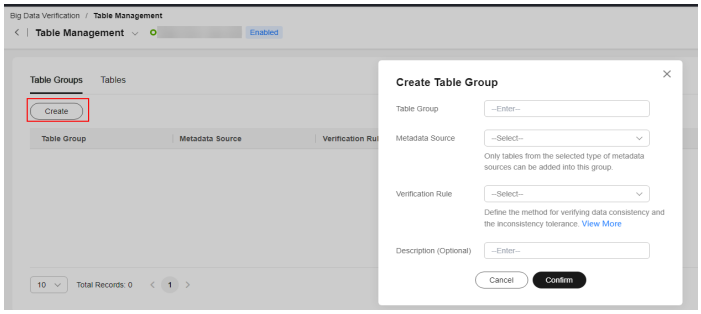
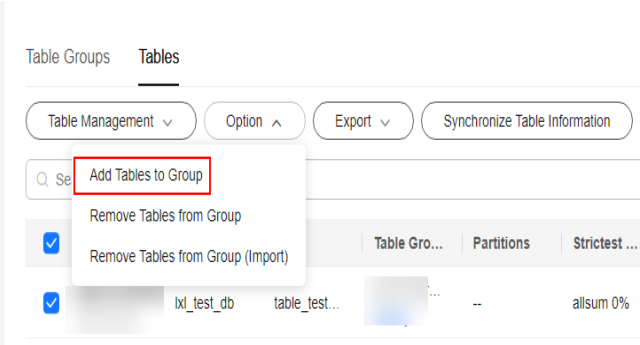


Table 7-19 Parameters for creating a table group

Parameter	Description
Table Group	User-defined
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see more information about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



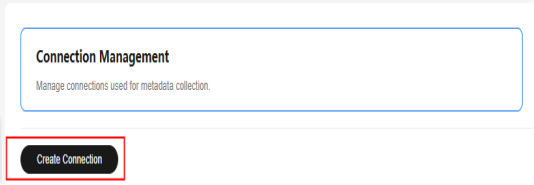
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

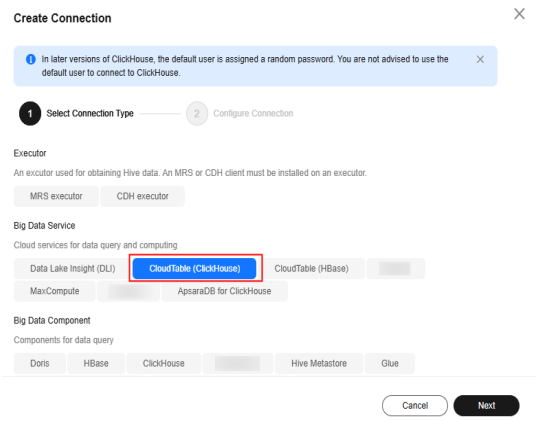
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-35 Creating a connection



Step 24 On the **Select Connection Type** page, select **CloudTable (ClickHouse)** and click **Next**.



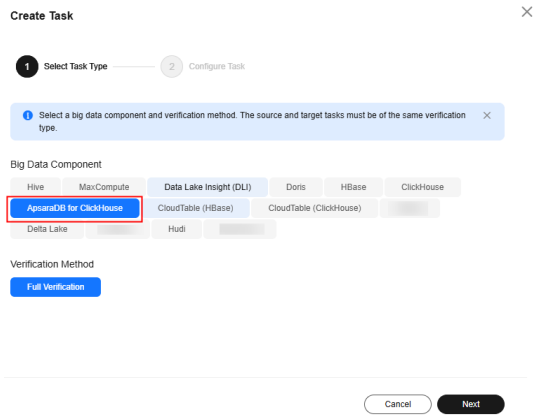
Step 25 Configure the **parameters for creating a connection to ClickHouse**, and click **Test**. If the test is successful, the connection is set up.

Table 7-20 Parameters for creating a ClickHouse connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **CloudTable-ClickHouse***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
CloudTable (ClickHouse) Credential	Select the credential you added to the MgC Agent for accessing the target CloudTable (ClickHouse) cluster in **step 6**.
Database URL	Enter the URL to access the CloudTable (ClickHouse) database. You can obtain the access address from the basic information of the ClickHouse cluster.

Step 26 Create a data verification task for the source Alibaba Cloud ApsaraDB for ClickHouse cluster, and execute the task. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

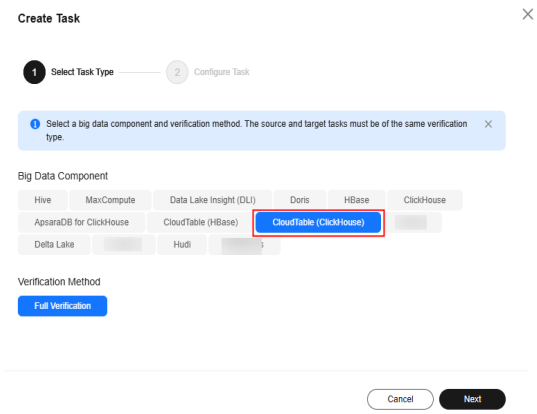
- On the **Select Task Type** page, choose **ApsaraDB for ClickHouse**.



- Select **Full Verification** for **Verification Method**.

Step 27 Create a data verification task for the CloudTable (ClickHouse) cluster, and execute the task. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **CloudTable (ClickHouse)**.



- Select **Full Verification** for **Verification Method**.

Step 28 Wait until the task executions enter a **Completed** status. On the **Verification Results** page, you can view and export the task execution results. For details, see **Viewing and Exporting Verification Results**.

----End

7.6 Verifying the Consistency of Data Migrated Between MRS Doris Clusters

This section describes how to use MgC to verify the consistency of data migrated between different versions of Huawei Cloud MRS Doris clusters.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

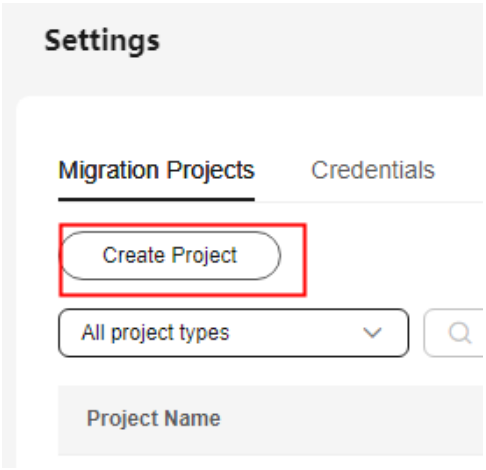
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

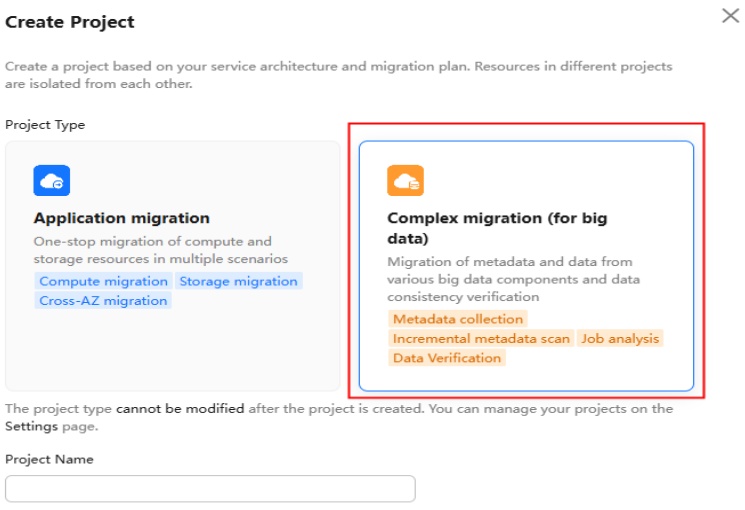
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-36 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-37 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source and target MRS Doris clusters to the MgC Agent. For more information, see **Adding Resource Credentials**.

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

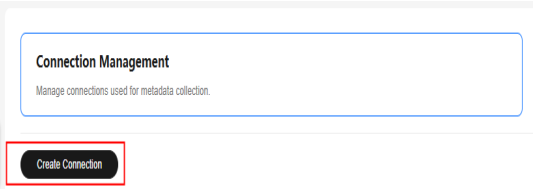
Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

> **CAUTION**
>
> Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

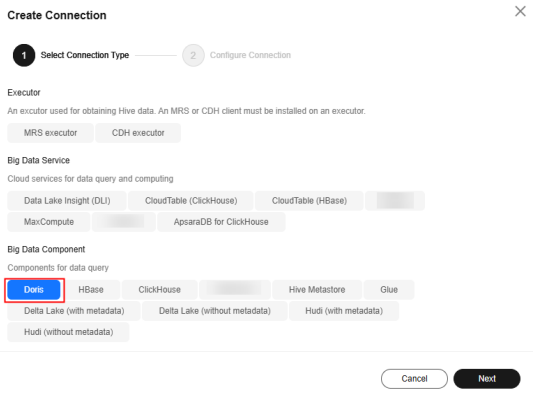
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-38 Creating a connection



Step 11 On the **Select Connection Type** page, select **Doris** and click **Next**.



Step 12 Configure the **parameters for creating a connection to Doris**, and click **Test**. If the test is successful, the connection is set up.

Table 7-21 Parameters for creating a connection to Doris

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Doris-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.

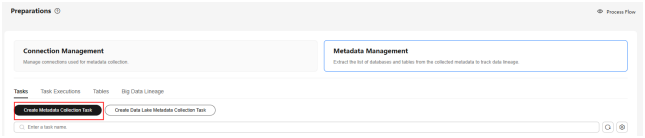
Parameter	Configuration
Doris Credential	Select the credential you added to the MgC Agent for accessing the source MRS Doris cluster in **step 6**.
Database IP Address	Enter the IP address of the Doris FE instance that is connected to the source Doris database. To obtain the IP address of a Doris FE instance, log in to the **MRS Manager** of the cluster and choose **Cluster** > **Services** > **Doris** > **Instances** to view the IP address of any FE instance.
Database Port	Enter the port for connecting to the source Doris database. The database connection port is the query connection port of the Doris FE. To obtain the port, you can log in to the MRS Manager, choose **Cluster** > **Services** > **Doris** > **Configurations**, and query the value of **query_port** of the Doris service.
Database Name	Enter the name of the source Doris database.

Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-39 Create Metadata Collection Task



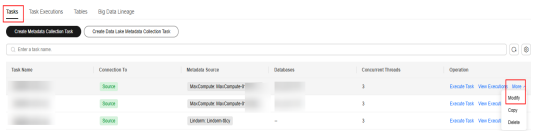
Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

Table 7-22 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also specify a name.
Metadata Connection	Select the connection created in **step 12**.
Databases	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

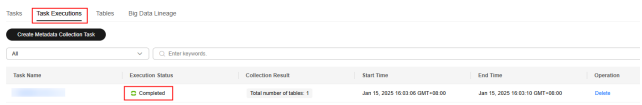
Figure 7-40 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-41 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

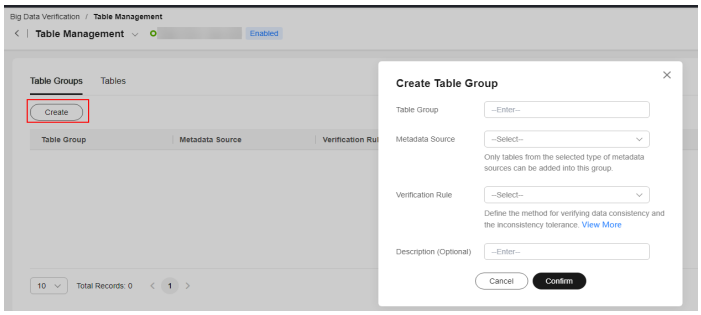
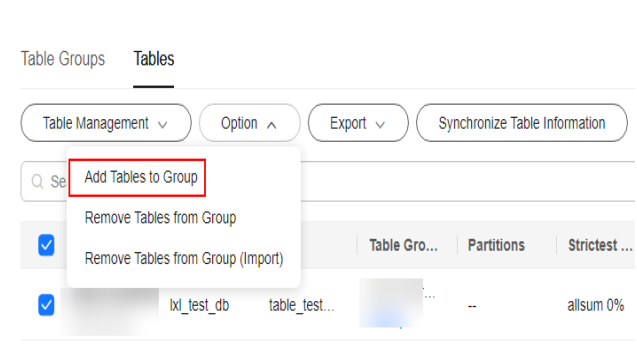


Table 7-23 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see the details about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



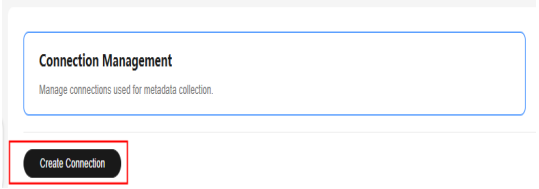
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

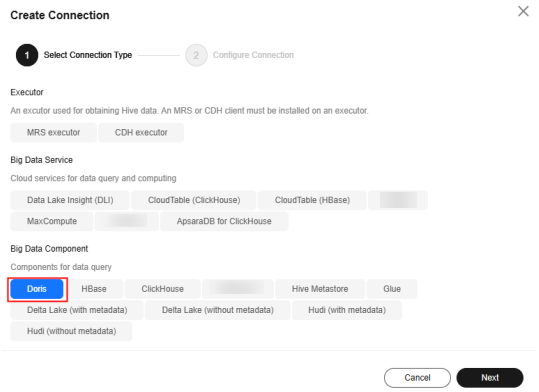
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-42 Creating a connection



Step 24 On the **Select Connection Type** page, select **Doris** and click **Next**.



Step 25 Set connection parameters based on **Table 7-24** and click **Test**. If the test is successful, the connection is set up.

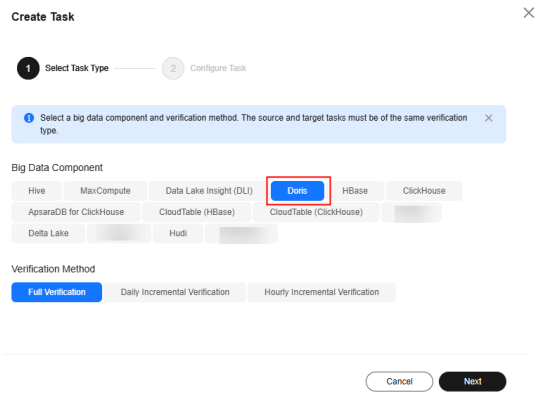
Table 7-24 Parameters for creating a connection to Doris

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **Doris-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Doris Credential	Select the credential you added to the MgC Agent for accessing the target MRS HBase cluster in **step 6**.
Database IP Address	Enter the IP address of the Doris FE instance that is connected to the target Doris database. To obtain the IP address of a Doris FE instance, log in to the **MRS Manager** of the cluster and choose **Cluster** > **Services** > **Doris** > **Instances** to view the IP address of any FE instance.

Parameter	Configuration
Database Port	Enter the port for connecting to the target Doris database. The database connection port is the query connection port of the Doris FE. To obtain the port, you can log in to MRS Manager, choose **Cluster** > **Services** > **Doris** > **Configurations**, and query the value of **query_port** of the Doris service.
Database Name	Enter the name of the MRS Doris database where the data to be verified is located.
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 26 On the MgC console, create a verification task for the source and target clusters, respectively, and execute the tasks. For details, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **Doris**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 27 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.7 Verifying the Consistency of Data Migrated Between MRS Hive Clusters or from CDH or EMR to MRS Hive

This section describes how to use MgC to verify the consistency of data migrated between Huawei Cloud MRS Hive clusters or migrated from self-built CDH or EMR clusters to Huawei Cloud MRS Hive clusters.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

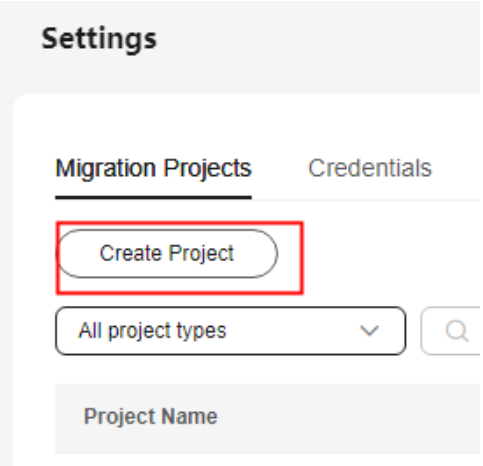
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

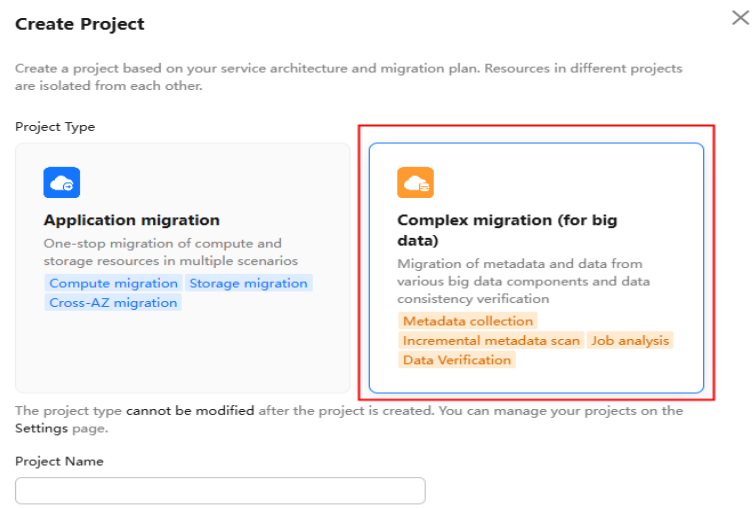
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-43 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-44 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source and target executors to the MgC Agent. For more information, see **Adding Resource Credentials**.

NOTICE

If the source MRS Hive cluster is secured (with Kerberos authentication enabled), add the Hive Metastore credential. You need to set **Type** to **Big Data** - **Hive Metastore** and **Authentication** to **Username/Key**. Upload the **core-site.xml**, **hivemetastore-site.xml**, **hive-site.xml**, **krb5.conf**, and **user.keytab** files. For details, see **How Do I Obtain the Hive Metastore Credential Files?**

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

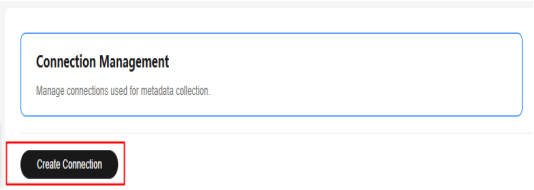
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

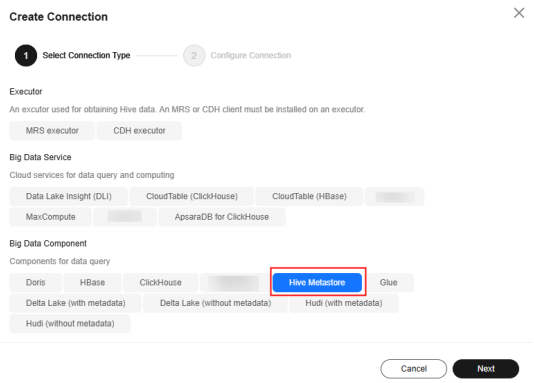
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-45 Creating a connection



Step 11 On the **Select Connection Type** page, select **Hive Metastore** and click **Next**.



Step 12 Configure the **parameters for creating a connection to Hive Metastore**, and click **Test**. If the test is successful, the connection is set up.

Table 7-25 Parameters for creating a connection to Hive Metastore

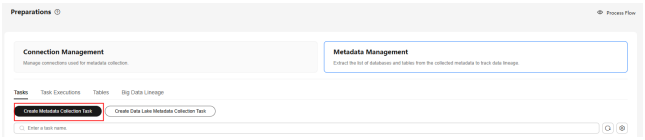
Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Hive-Metastore-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.

Parameter	Configuration
Secure Connection	Choose whether to enable secure connection. • If Hive Metastore is deployed in an unsecured cluster, do not enable secure connection. • If Hive Metastore is deployed in a secured cluster, enable secure connection and provide access credentials. Select the source Hive Metastore credential added to the MgC Agent in **step 6**.
Hive Version	Select the source Hive version.
Hive Metastore IP Address	Enter the IP address for connecting to the Hive Metastore node.
Hive Metastore Thrift Port	Enter the port for connecting to the Hive Metastore Thrift service. The default port is **9083**.
Connect to Metadata Database	During an incremental data verification, querying with Hive Metastore on more than 30,000 partitions may lead to a memory overflow (OOM) since all partition information is loaded into memory. Connecting to the MySQL metadata database can effectively prevent this issue. • If you disable this option, the system queries the information of Hive tables and partitions using Hive Metastore. • If you enable this option, configure the MySQL database information. The system will query the information of Hive tables and partitions through the MySQL database. You need to set the following parameters: – **Metadata Database Type**: Only MySQL is supported. – **MySQL Credential**: Select the credential for accessing the MySQL database. You need to add the credential to the MgC Agent and synchronize it to MgC. For more information, see **Adding Resource Credentials**. – **MySQL Node IP Address**: Enter the IP address of the MySQL database server. – **MySQL Port**: Enter the port of the MySQL database service. – **Database Name**: Enter the name of the database that stores the Hive table metadata. **NOTE** Ensure that the entered MySQL credential, node IP address, service port, and database name match the MySQL database used by Hive. Otherwise, data verification will fail.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-46 Create Metadata Collection Task



Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

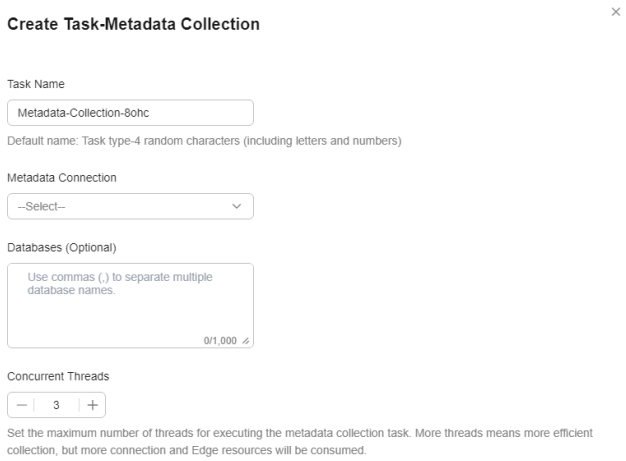
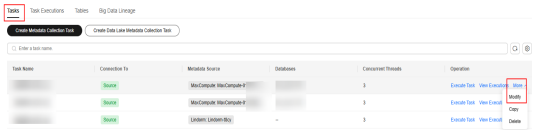


Table 7-26 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also specify a name.
Metadata Connection	Select the connection created in **step 12**.
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

Figure 7-47 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-48 Managing task executions

Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

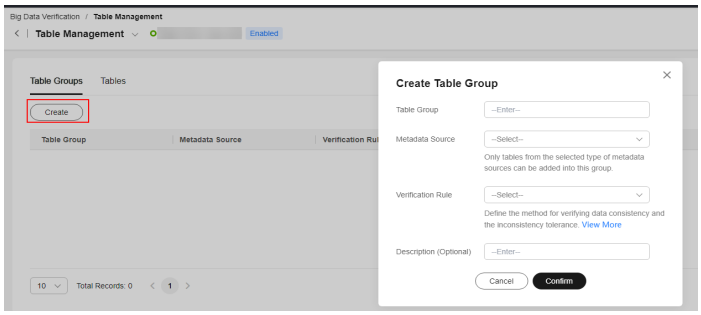
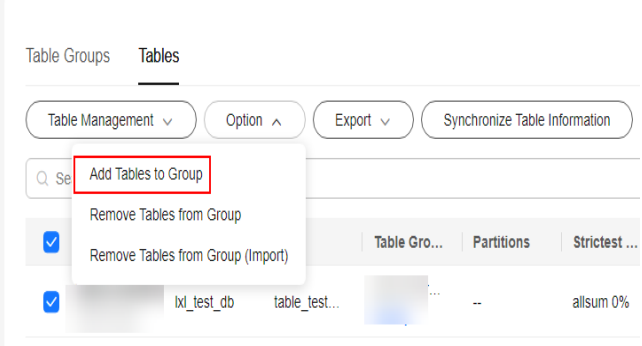


Table 7-27 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see the details about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



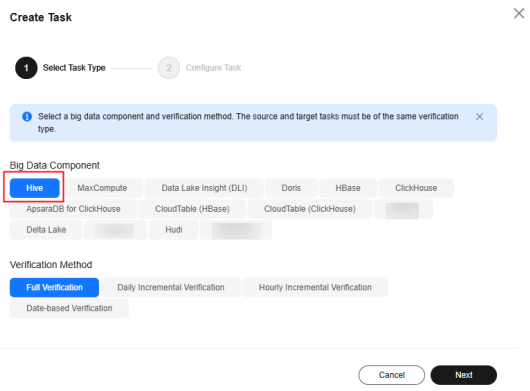
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

Step 22 Create a connection to the source and target executors separately. For details, see **Creating an Executor Connection**. Select the source and target executor credentials added to the MgC Agent in **step 6**.

Step 23 Create a data verification task for the source and target Hive clusters, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **Hive**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 24 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.8 Verifying the Consistency of Data Migrated from MaxCompute to MRS Hive

This section describes how to use MgC to verify the consistency of data migrated from Alibaba Cloud MaxCompute to Huawei Cloud MRS Hive.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

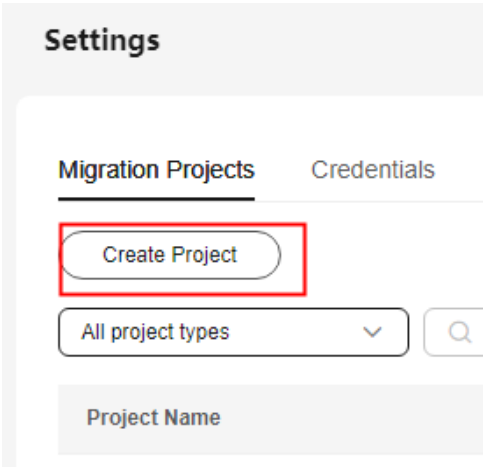
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

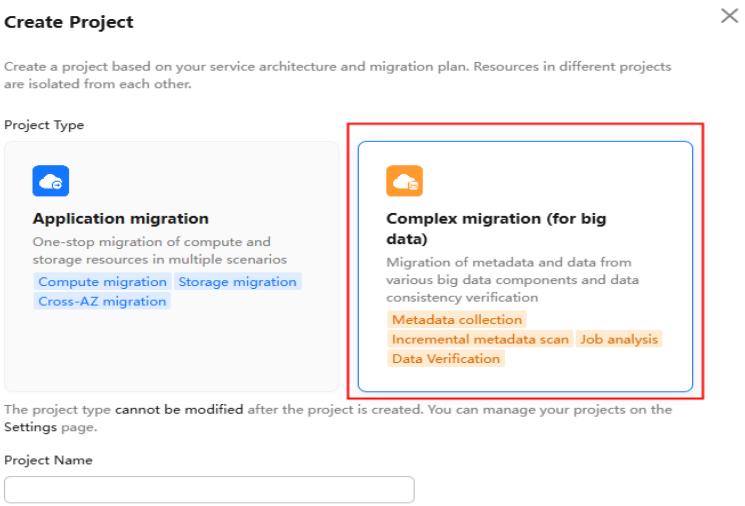
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-49 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-50 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the AK/SK pair for accessing MaxCompute and the username/passwords pairs for accessing Hive Metastore and MRS executor to the MgC Agent. For more information, see **Adding Resource Credentials**.

- For details about how to obtain an AK/SK pair for accessing MaxCompute, see **Viewing the Information About AccessKey Pairs of a RAM User**.
- For details about how to obtain the Hive Metastore credential files, see **How Do I Obtain the Hive Metastore Credential Files?**

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

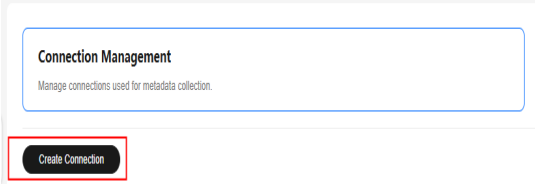
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

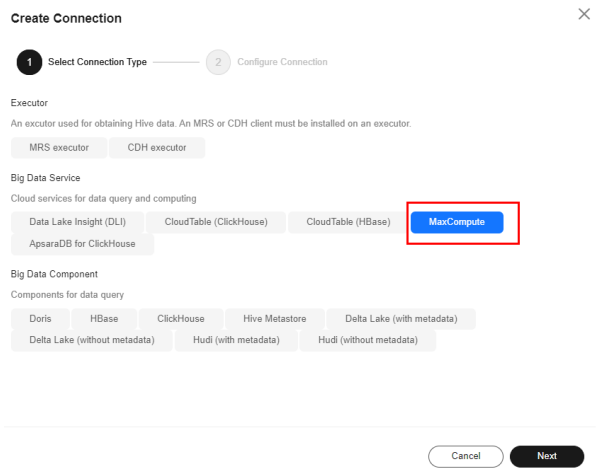
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-51 Creating a connection



Step 11 On the **Select Connection Type** page, select **MaxCompute** and click **Next**.



Step 12 Configure the **parameters for creating a connection to MaxCompute**, and click **Test**. If the test is successful, the connection is set up.

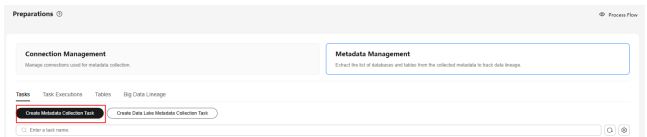
Table 7-28 Parameters for creating a connection to MaxCompute

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **MaxCompute-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Alibaba Cloud Credential	Select the MaxCompute credential added to the MgC Agent in **step 6**.
MaxCompute Project	Enter the name of your MaxCompute project. You can obtain the project name from the MaxCompute console.
Endpoint	Enter the endpoint of the region where the MaxCompute project is located. For details about the MaxCompute endpoints in different regions, see **MaxCompute Endpoints**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-52 Create Metadata Collection Task



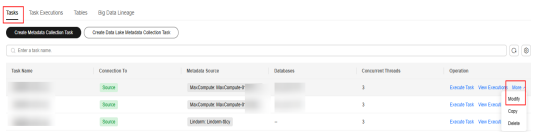
Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

Table 7-29 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also specify a name.
Metadata Connection	Select the connection created in **step 12**.
Databases	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. **NOTICE** This parameter is mandatory only if a MaxCompute metadata connection is selected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

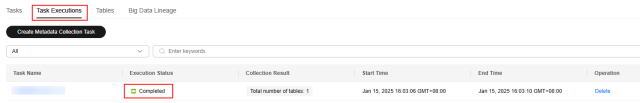
Figure 7-53 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-54 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

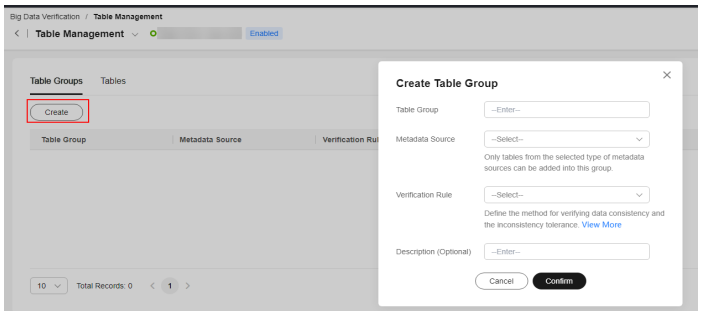
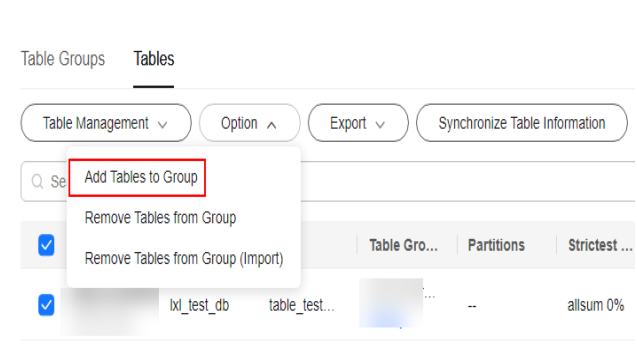


Table 7-30 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see the details about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



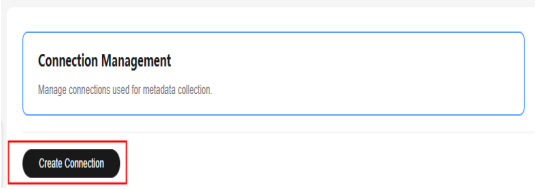
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

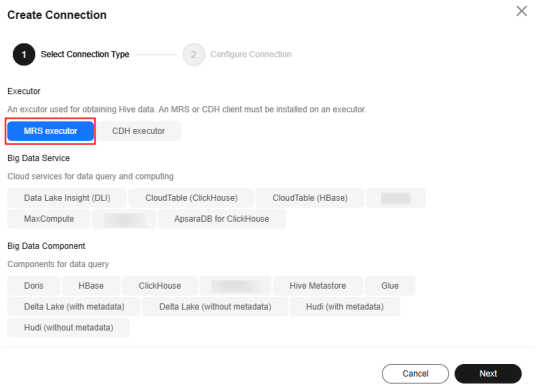
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-55 Creating a connection



Step 24 On the **Select Connection Type** page, select **MRS executor** and click **Next**.



Step 25 Set connection parameters based on **Table 7-31** and click **Test**. If the test is successful, the connection is set up.

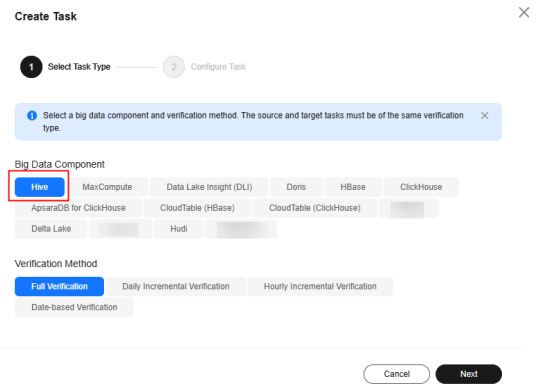
Table 7-31 Parameters for creating an executor connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is *Executor type*-4 random characters (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Executor Credential	Select the MRS executor credential added to the MgC Agent in **step 6**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor. The default port is **22**.
Installation Directory	Enter the installation directory of the MRS client. That is, the directory where **./install.sh** is installed.

Parameter	Configuration
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 26 On the MgC console, create a verification task for the source and target Hive clusters, respectively, and execute the tasks. For details, see **Creating and Executing Verification Tasks**. Select the table group created in **step 20** and the MRS executor connection created in **step 25**.

- On the **Select Task Type** page, choose **Hive**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 27 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.9 Verifying the Consistency of Data Migrated Between MRS HBase Clusters

This section describes how to use MgC to verify the consistency of data migrated between different versions of Huawei Cloud MRS HBase clusters.

Preparations

- Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in to the tool. For details, see **Installing the MgC Agent on Linux**.
- Add the mappings between the hostnames and IP addresses of all nodes in the source and target clusters to the **/etc/hosts** file on the server where the MgC Agent is installed.
  a. On the server where the MgC Agent is installed, open the **/etc/hosts** file.
  b. In the **/etc/hosts** file, add a line for each node in the source and target clusters in the following format:
  IP address Hostname
  For example, if a node uses the IP address **192.168.1.1** and has the hostname **source-node-01**, add the following information:
  192.168.1.1 source-node-01
  c. After all mappings are added, save and close the **/etc/hosts** file.
  d. Ping a hostname to check whether it can be resolved successfully. For example:
  ping source-node-01

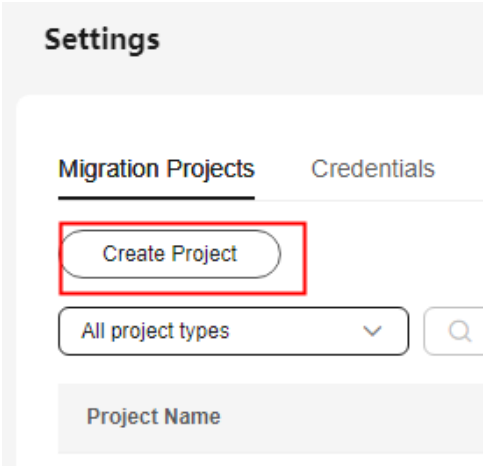
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

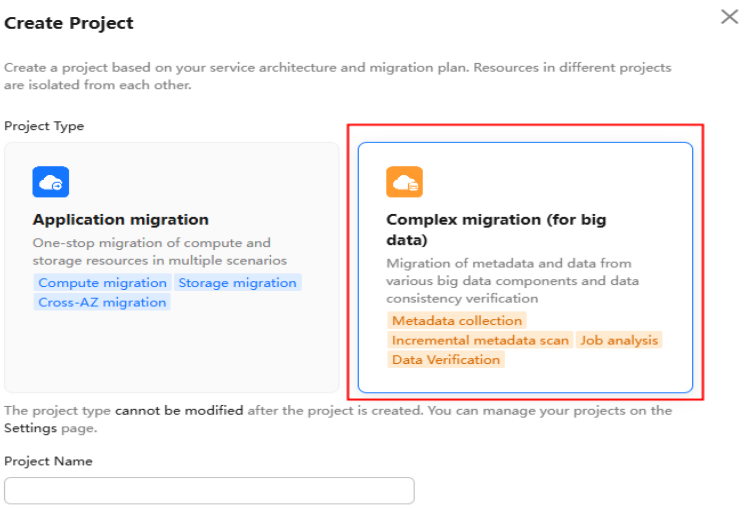
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-56 Creating a project



Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-57 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/key pairs for accessing the source and target MRS HBase clusters to the MgC Agent. For more information, see **Adding Resource Credentials**. Enter the username for logging in to the HBase client in the **Username** box and upload the configuration files for **Key**:

- For an unsecured cluster (with Kerberos authentication disabled), you need to upload five configuration files: **core-site.xml**, **hdfs-site.xml**, **yarn-site.xml**, **mapred-site.xml**, and **hbase-site.xml**.
- For a secured cluster (with Kerberos authentication enabled), upload seven files: **core-site.xml**, **hdfs-site.xml**, **yarn-site.xml**, **krb5.conf**, **user.keytab**, **mapred-site.xml**, and **hbase-site.xml**. The **krb5.conf** and **user.keytab** files contain the credentials of the cluster users. You can perform the following steps to obtain the two files.
  a. Log in to **FusionInsight Manager**, and choose **System** > **Permission** > **User**.
  b. Select **developuser** and choose **More** > **Download Authentication Credential** to download the authentication credential files.
  c. Decompress the downloaded file to obtain the **user.keytab** and **krb5.conf** files.

Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

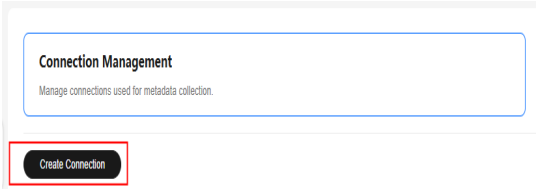
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

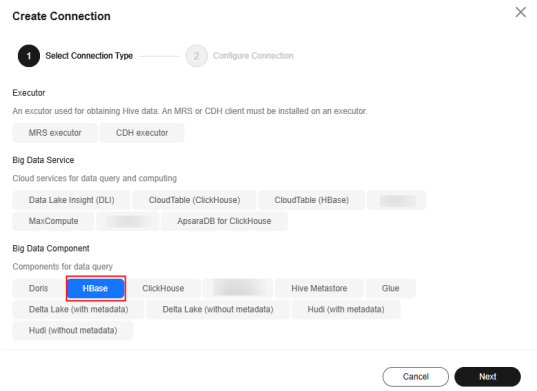
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-58 Creating a connection



Step 11 On the **Select Connection Type** page, select **HBase** and click **Next**.



Step 12 Set connection parameters based on **Table 7-32** and click **Test**. If the test is successful, the connection is set up.

Table 7-32 Parameters for creating a connection to HBase

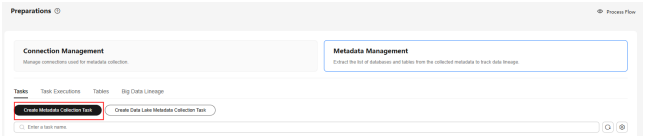
Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **HBase-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
HBase Credential	Select the credential you added to the MgC Agent for accessing the source MRS HBase cluster in **step 6**.
Secured Cluster	Choose whether the cluster is secured.
ZooKeeper IP Address	Enter the IP address for connecting to the ZooKeeper node. You can enter the public or private IP address of the ZooKeeper node.
ZooKeeper Port	Enter the port for connecting to the ZooKeeper node.
HBase Version	Select the HBase version.

Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Metadata Collection Task**.

Figure 7-59 Create Metadata Collection Task



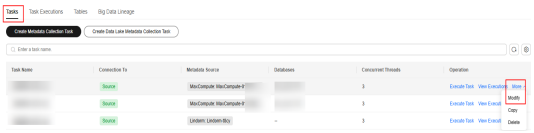
Step 15 Configure the **parameters for creating a metadata collection task** and click **Confirm**.

Table 7-33 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also specify a name.
Metadata Connection	Select the connection created in **step 12**.
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. Use commas (,) to separate the database names. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

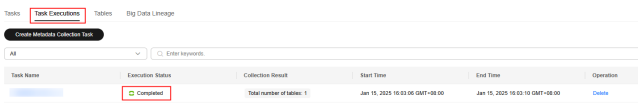
Figure 7-60 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-61 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

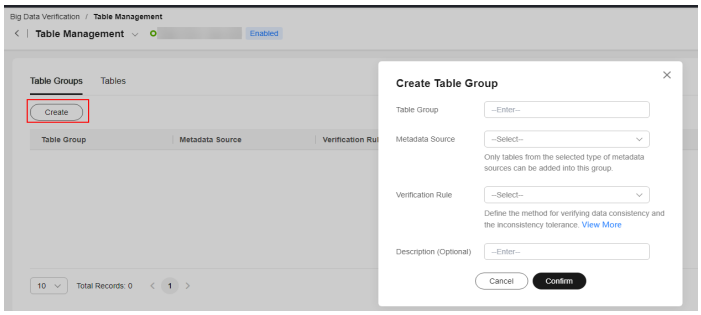
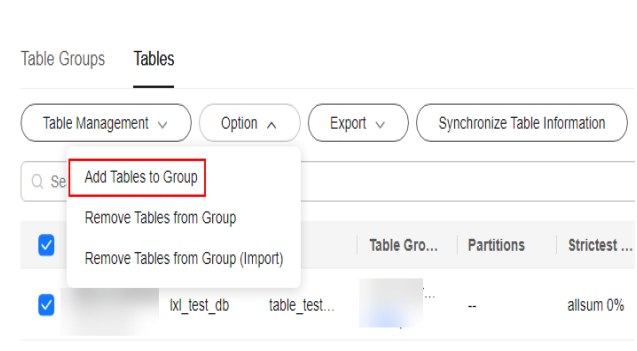


Table 7-34 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see the details about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



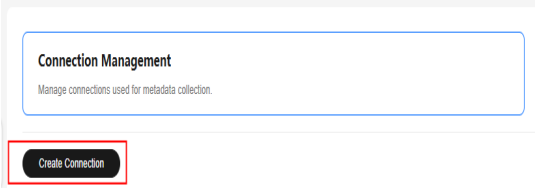
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

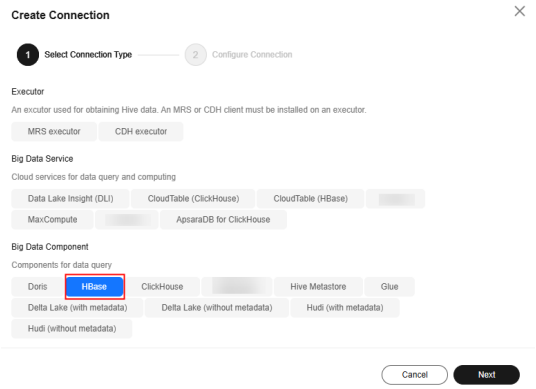
Step 22 In the **Features** area, click **Migration Preparations**.

Step 23 Choose **Connection Management** and click **Create Connection**.

Figure 7-62 Creating a connection



Step 24 On the **Select Connection Type** page, select **HBase** and click **Next**.



Step 25 Set connection parameters based on **Table 7-35** and click **Test**. If the test is successful, the connection is set up.

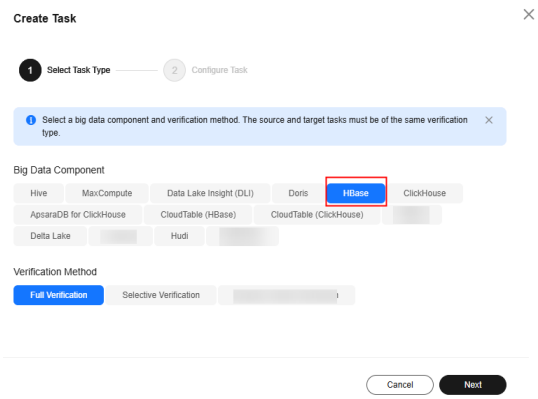
Table 7-35 Parameters for creating an HBase connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **HBase-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
HBase Credential	Select the credential you added to the MgC Agent for accessing the target MRS HBase cluster in **step 6**.
Secured Cluster	Choose whether the cluster is secured.
ZooKeeper IP Address	Enter the IP address for connecting to the ZooKeeper node. You can enter the public or private IP address of the ZooKeeper node.
ZooKeeper Port	Enter the port for connecting to the ZooKeeper node. The default value is **2181**.
HBase Version	Select the HBase version.

Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 26 On the MgC console, create a verification task for the source and target HBase clusters, respectively, and execute the tasks. For details, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **HBase**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 27 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.10 Verifying the Consistency of Data Migrated from Delta Lake (with Metadata) to MRS Delta Lake

This section describes how to use MgC to verify the consistency of data migrated from self-built Delta Lake clusters to Huawei Cloud MRS Delta Lake clusters.

NOTICE

For Delta Lake clusters that have metadata storage, the metadata can be collected through data lake metadata collection tasks.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

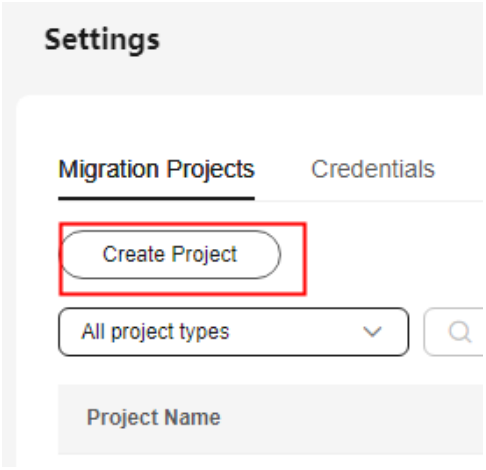
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

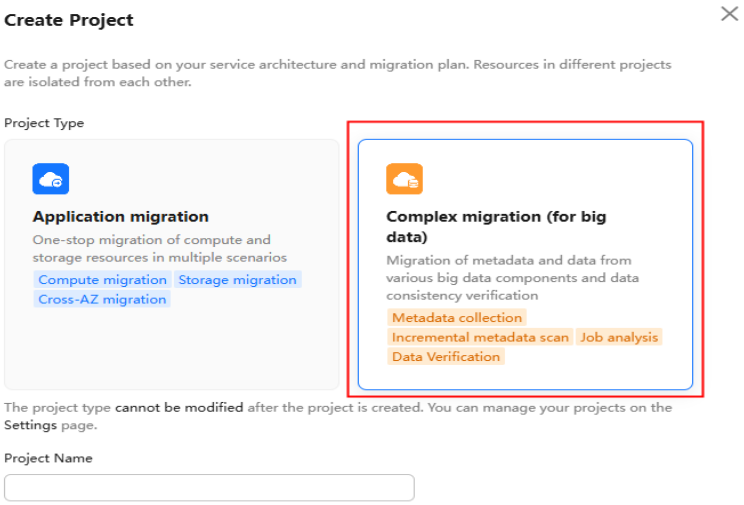
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-63 Creating a project



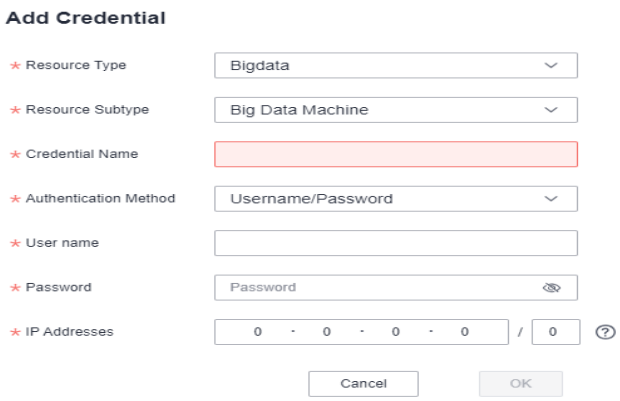
Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-64 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source Delta Lake executor and the target MRS Delta Lake executor to the MgC Agent. For more information, see **Adding Resource Credentials**.



Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

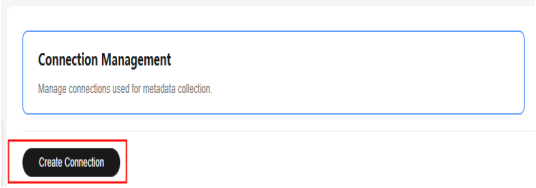
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

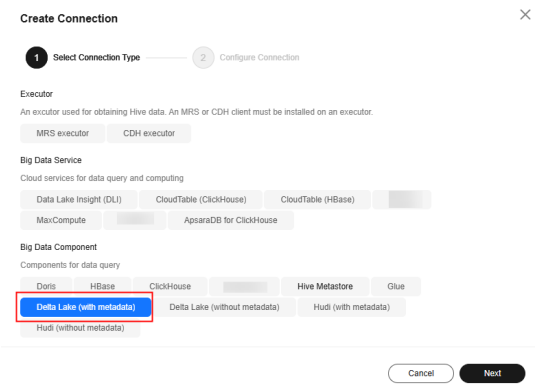
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-65 Creating a connection



Step 11 Select **Delta Lake (with metadata)** and click **Next**.



Step 12 Set connection parameters based on **Table 7-36** and click **Test**. If the test is successful, the connection is set up.

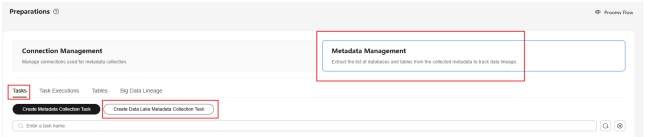
Table 7-36 Parameters for creating a connection to Delta Lake (with metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Delta-Lake-with-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Executor Credential	Select the source Delta Lake executor credential added to the MgC Agent in **step 6**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor. The default port is **22**.
Spark Client Directory	Enter the absolute path of the bin directory on the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file, for example, **/opt/bigdata/client/bigdata_env**. If this field is not left blank, the environment variable file is automatically sourced before commands are executed.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the folder. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Data Lake Metadata Collection Task**.

Figure 7-66 Create Data Lake Metadata Collection Task



Step 15 Set a data lake metadata collection task based on **Table 7-37** and click **Confirm**.

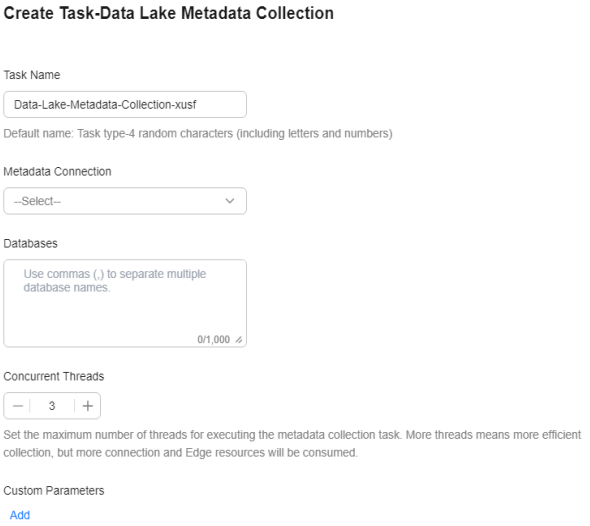


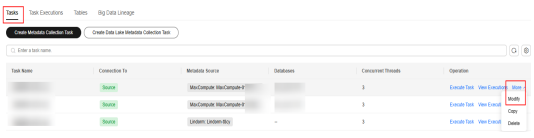
Table 7-37 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Data-Lake-Metadata-Collection-Task-***4 random characters* (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.
Databases	Enter the names of the databases whose metadata needs to be collected. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.

Parameter	Configuration
Custom Parameters	You can customize parameters to specify the tables and partitions to collect or set criteria to filter tables and partitions. • If the metadata source is Alibaba Cloud EMR, add the following parameter: – **Parameter**: **conf** – **Value**: **spark.sql.catalogImplementation=hive** • If the source is Alibaba Cloud EMR Delta Lake 2.2 and is accessed through Delta Lake 2.3 dependencies, add the following parameter: – **Parameter**: **master** – **Value**: **local** • If you are creating a verification task for an Alibaba Cloud EMR Delta Lake 2.1.0 cluster that uses Spark 2.4.8, add the following parameter: – **Parameter**: **mgc.delta.spark.version** – **Value**: **2** • If the source is Alibaba Cloud EMR and is configured with Spark 3 to process Delta Lake data, add the following parameter: – **Parameter**: **jars** – **Value**: **'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'** **CAUTION** Replace the parameter values with the actual environment directory and Delta Lake version.

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

Figure 7-67 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-68 Managing task executions

Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

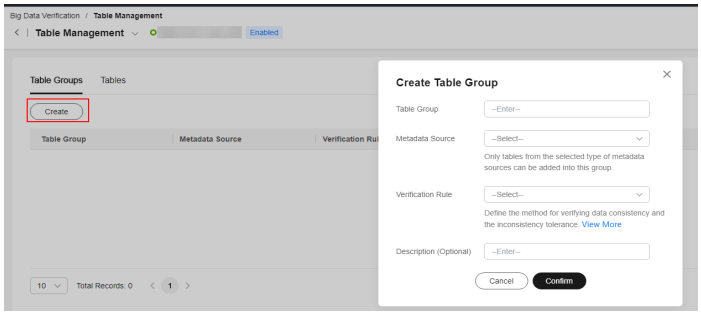
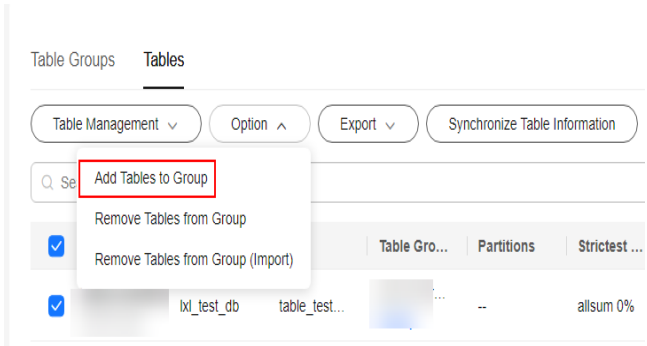


Table 7-38 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. MgC provides multiple verification rules for you to choose. For details about these rules, click **View More**.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



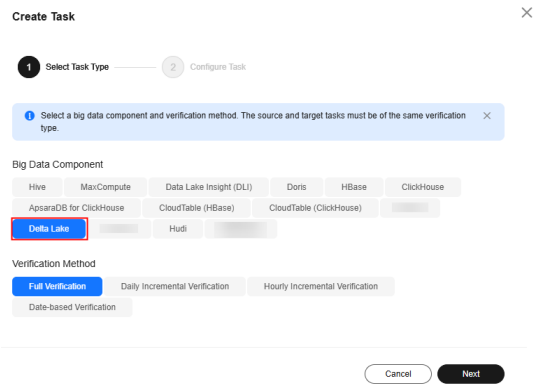
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

Step 22 Create a connection to the source and target executors separately. For details, see **Creating an Executor Connection**. Select the source and target executor credentials added to the MgC Agent in **step 6**.

Step 23 Create a data verification task for the source and target Delta Lake clusters, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **Delta Lake**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 24 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.11 Verifying the Consistency of Data Migrated from Delta Lake (without Metadata) to MRS Delta Lake

This section describes how to use MgC to verify the consistency of data migrated from self-built Delta Lake clusters to Huawei Cloud MRS Delta Lake clusters.

NOTICE

For Delta Lake clusters without metadata storage, you need to import the metadata to MgC.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

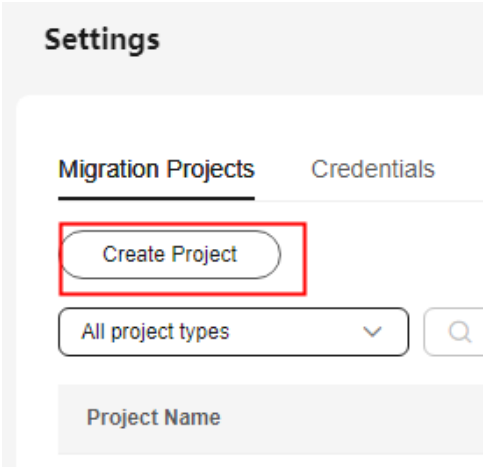
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

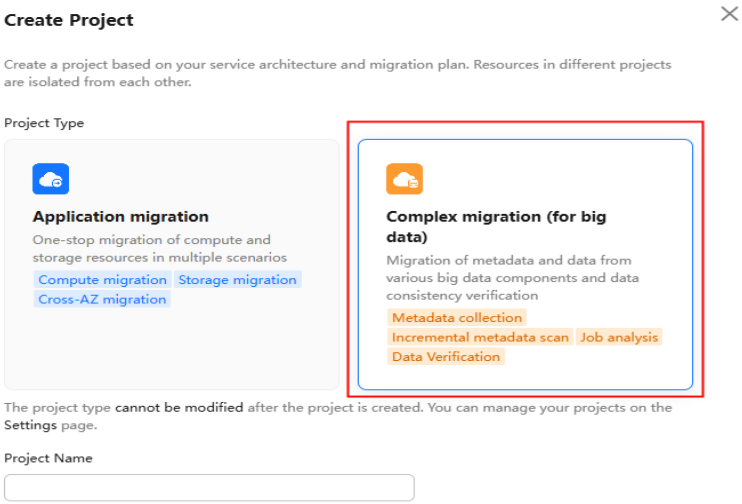
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-69 Creating a project



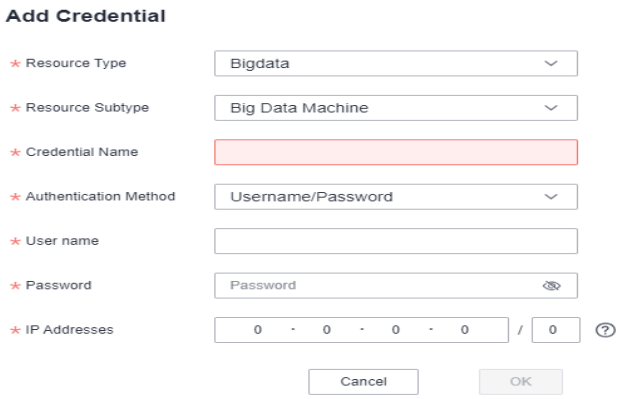
Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-70 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source Delta Lake executor and the target MRS Delta Lake executor to the MgC Agent. For more information, see **Adding Resource Credentials**.



Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

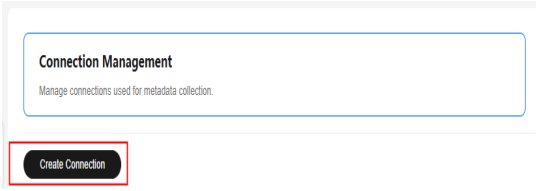
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

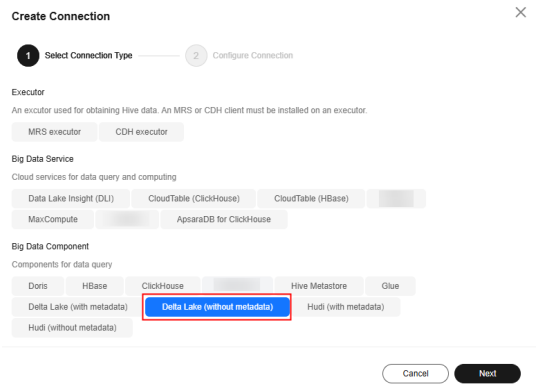
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-71 Creating a connection



Step 11 Select **Delta Lake (without metadata)** and click **Next**.



Step 12 Set connection parameters based on **Table 7-39** and click **Test**. If the test is successful, the connection is set up.

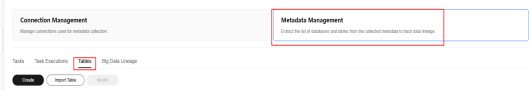
Table 7-39 Parameters for creating a connection to Delta Lake (without metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Delta-Lake-without-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Executor Credential	Select the source Delta Lake executor credential added to the MgC Agent in **step 6**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor. The default port is **22**.
Spark Client Directory	Enter the absolute path of the bin directory on the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file, for example, **/opt/bigdata/client/bigdata_env**. If this field is not left blank, the environment variable file is automatically sourced before commands are executed.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the folder. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

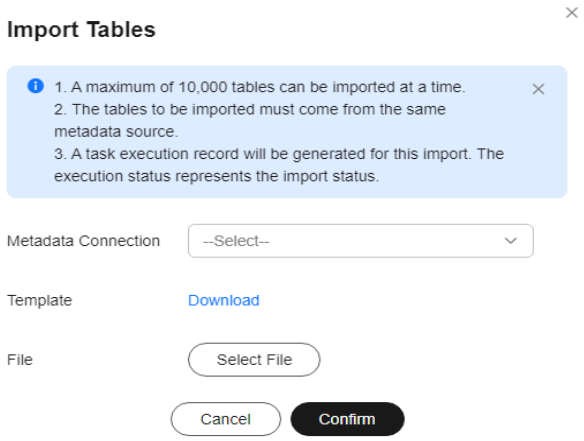
Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click the **Tables** tab.

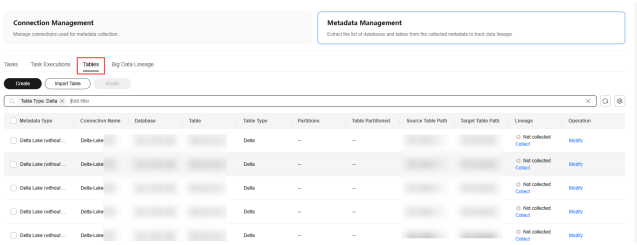


Step 15 On the displayed page, click **Import**.



Step 16 Click **Download** to download the import template to the local PC. Open the template, fill in table information, and save the template.

Step 17 In the **Import Tables** dialog box, click **Select File**, choose the filled template file, and click **Confirm**. After the import is complete, you can view the imported tables on **Tables** tab.



Step 18 In the **Features** area, click **Table Management**.

Step 19 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

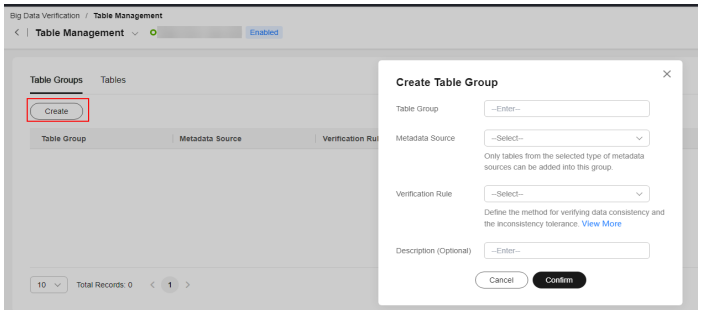
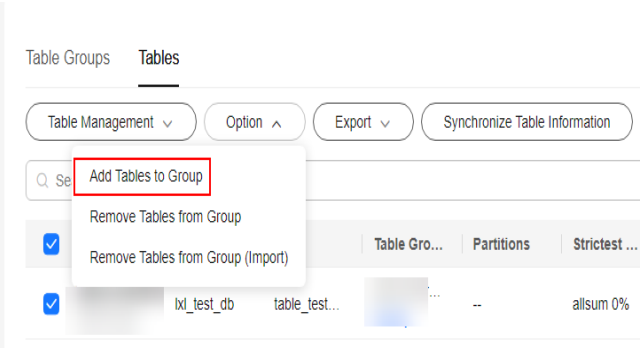


Table 7-40 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.

Parameter	Description
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. MgC provides multiple verification rules for you to choose. For details about these rules, click **View More**.
Description (Optional)	Enter a description to identify the table group.

Step 20 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



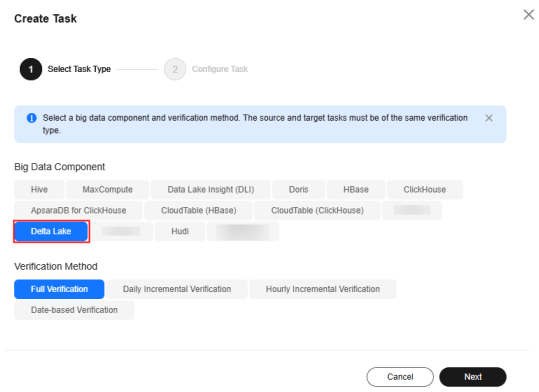
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

Step 21 Create a connection to the source and target executors separately. For details, see **Creating an Executor Connection**. Select the source and target executor credentials added to the MgC Agent in **step 6**.

Step 22 Create a data verification task for the source and target Delta Lake clusters, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 19**.

- On the **Select Task Type** page, choose **Delta Lake**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 23 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.12 Verifying the Consistency of Data Migrated from Hudi (with Metadata) to MRS Hudi

This section describes how to use MgC to verify the consistency of data migrated from self-built Hudi clusters to Huawei Cloud MRS Hudi clusters.

NOTICE

For Hudi clusters that have metadata storage, the metadata can be collected through data lake metadata collection tasks.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

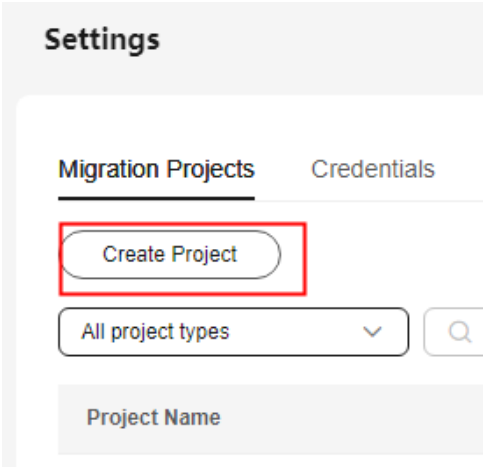
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

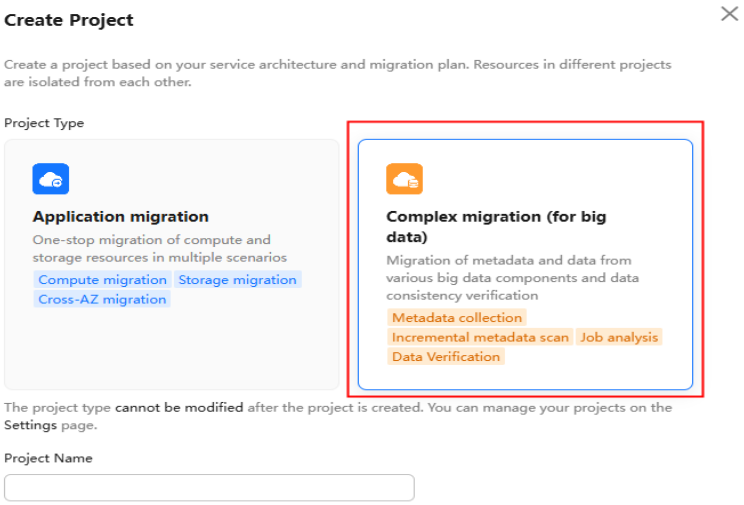
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-72 Creating a project



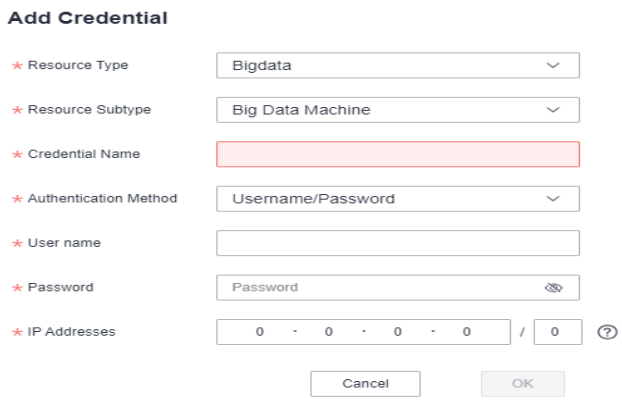
Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-73 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source Hudi executor and the target MRS Hudi executor to the MgC Agent. For more information, see **Adding Resource Credentials**.



Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

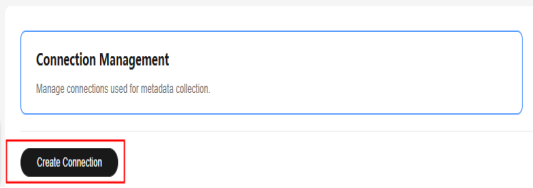
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

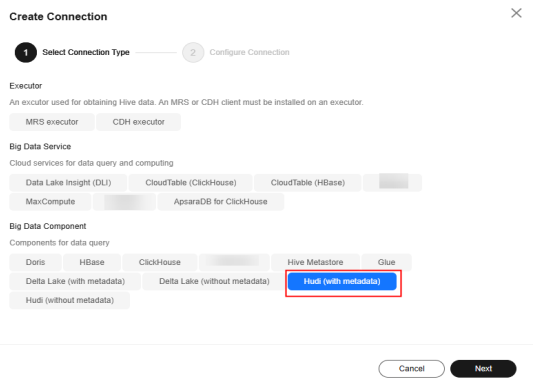
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-74 Creating a connection



Step 11 Select **Hudi (with metadata)** and click **Next**.



Step 12 Configure the connection parameters based on **Table 7-41** and click **Test**. If the test is successful, the connection is set up.

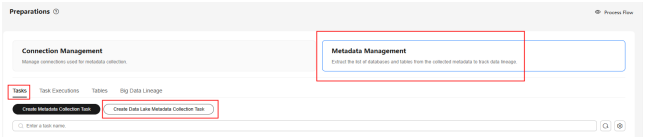
Table 7-41 Parameters for creating a connection to Hudi (with metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Hudi-with-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Executor Credential	Select the source Hudi executor credential added to the MgC Agent in **step 6**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor. The default port is **22**.
Spark Client Directory	Enter the absolute path of the bin directory on the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file, for example, **/opt/bigdata/client/bigdata_env**. If this field is not left blank, the environment variable file is automatically sourced before commands are executed.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click **Create Data Lake Metadata Collection Task**.

Figure 7-75 Create Data Lake Metadata Collection Task



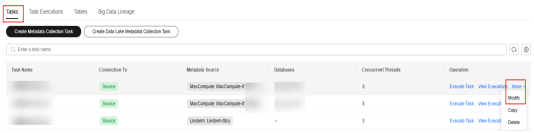
Step 15 Set a data lake metadata collection task based on **Table 7-42** and click **Confirm**.

Table 7-42 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Data-Lake-Metadata-Collection-Task-*4 random characters*** (including letters and numbers). You can also customize a name.
Metadata Connection	Select the connection created in **step 12**.
Databases (Optional)	Enter the names of the databases whose metadata needs to be collected. If no database name is specified, the metadata of all databases is collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. Configuring more concurrent threads means more efficient collection, but more connection and MgC Agent resources will be consumed.
Custom Parameters	You can customize parameters to specify the tables and partitions to collect or set criteria to filter tables and partitions. If you want to collect metadata from Alibaba Cloud EMR, add the following parameter: • **Parameter**: **conf** • **Value**: **spark.sql.catalogImplementation=hive**

Step 16 Under **Tasks**, review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

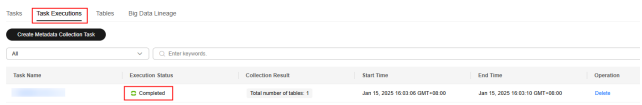
Figure 7-76 Managing a metadata collection task



Step 17 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 18 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Figure 7-77 Managing task executions



Step 19 In the **Features** area, click **Table Management**.

Step 20 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

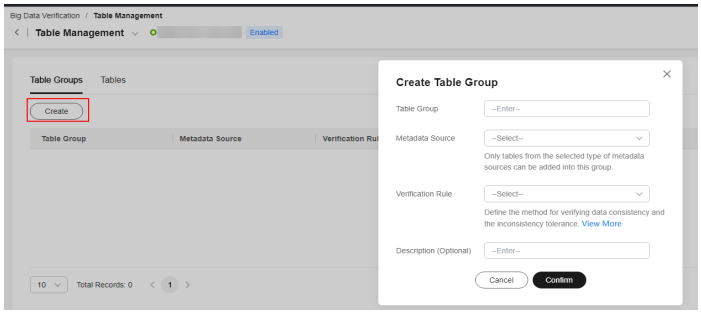
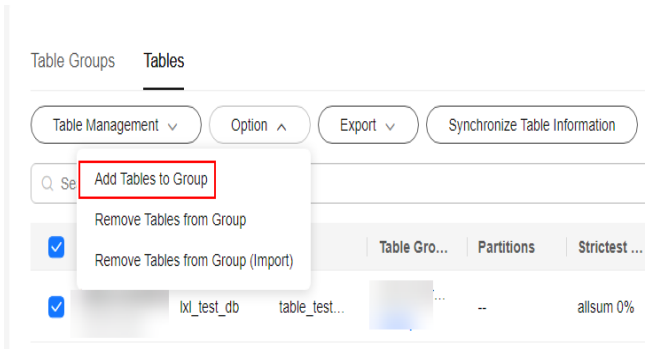


Table 7-43 Parameters for creating a table group

Parameter	Description
Table Group	User-defined
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. You can **View More** to see more information about the verification rules provided by MgC.
Description (Optional)	Enter a description to identify the table group.

Step 21 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



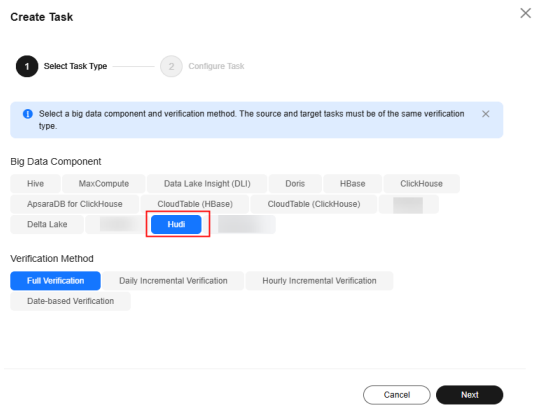
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

Step 22 Create a connection to the source and target executors separately. For details, see **Creating an Executor Connection**. Select the source and target executor credentials added to the MgC Agent in **step 6**.

Step 23 Create a data verification task for the source and target Hudi clusters, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 20**.

- On the **Select Task Type** page, choose **Hudi**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 24 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

7.13 Verifying the Consistency of Data Migrated from Hudi (Without Metadata) to MRS Hudi

This section describes how to use MgC to verify the consistency of data migrated from self-built Hudi clusters to Huawei Cloud MRS Hudi clusters.

NOTICE

For Hudi clusters without metadata storage, you need to import the metadata to MgC.

Preparations

Install the MgC Agent, a tool used for data verification, in the source intranet environment and log in. For details, see **Installing the MgC Agent on Linux**.

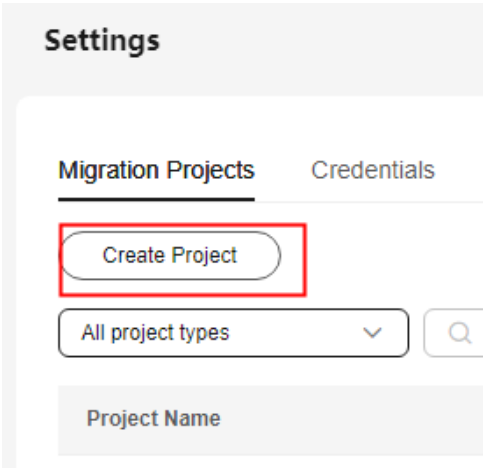
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Other** > **Settings**.

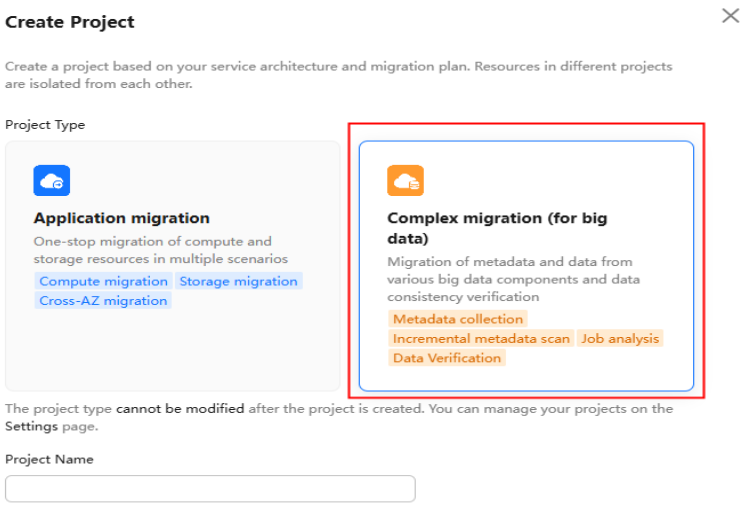
Step 3 Under **Migration Projects**, click **Create Project**.

Figure 7-78 Creating a project



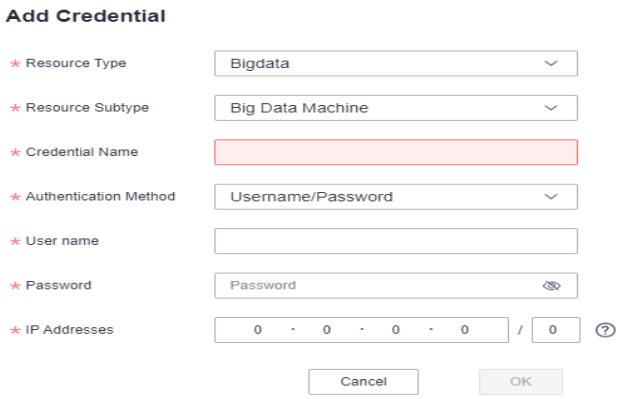
Step 4 Set **Project Type** to **Complex migration (for big data)**, enter a project name, and click **Create**.

Figure 7-79 Creating a big data migration project



Step 5 Connect the MgC Agent to MgC. For more information, see **Connecting the MgC Agent to MgC**.

Step 6 After the connection is successful, add the username/password pairs for accessing the source Hudi executor and the target MRS Hudi executor to the MgC Agent. For more information, see **Adding Resource Credentials**.



Step 7 In the navigation pane, choose **Migrate** > **Big Data Verification**. In the navigation pane, under **Project**, select the project created in step 4.

Step 8 If you are performing a big data verification with MgC for the first time, select your MgC Agent to enable this feature. Click **Select MgC Agent**. In the displayed dialog box, select the MgC Agent you connected to MgC from the drop-down list.

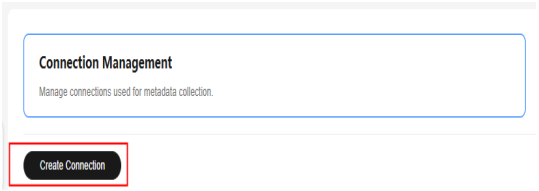
CAUTION

Ensure that the selected MgC Agent is always **Online** and **Enabled** before your verification is complete.

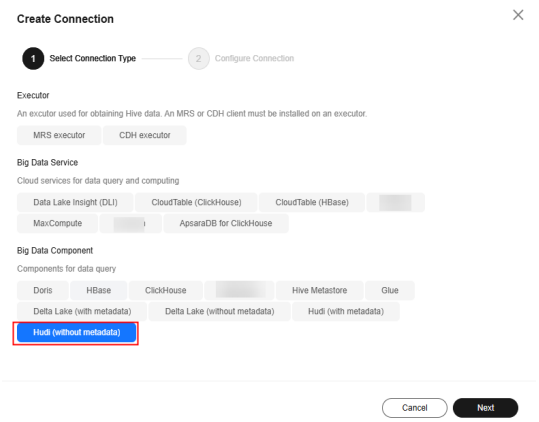
Step 9 In the **Features** area, click **Migration Preparations**.

Step 10 Choose **Connection Management** and click **Create Connection**.

Figure 7-80 Creating a connection



Step 11 Select **Hudi (without metadata)** and click **Next**.



Step 12 Configure the connection parameters based on **Table 7-44** and click **Test**. If the test is successful, the connection is set up.

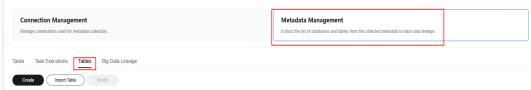
Table 7-44 Parameters for creating a connection to Hudi (without metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Hudi-without-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
MgC Agent	Select the MgC Agent connected to MgC in **step 5**.
Executor Credential	Select the source Hudi executor credential added to the MgC Agent in **step 6**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor. The default port is **22**.
Spark Client Directory	Enter the absolute path of the bin directory on the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file, for example, **/opt/bigdata/client/bigdata_env**. If this field is not left blank, the environment variable file is automatically sourced before commands are executed.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

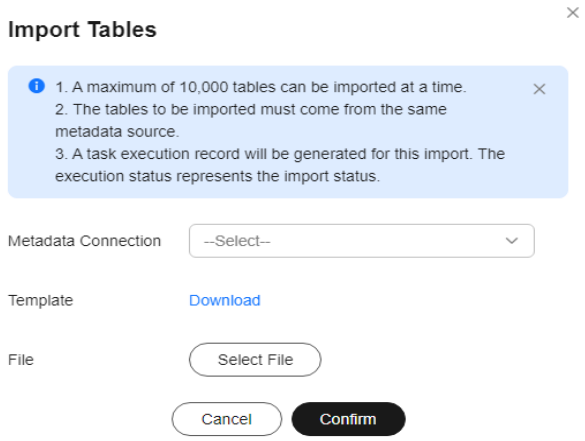
Parameter	Configuration
Collect Usage Metrics	This parameter is optional. If this option is enabled, usage metrics for your big data resources will be collected during the execution of tasks created using this connection. The collected information is used to generate reports on the MgC console and for performance optimization. **NOTICE** Before using this function, ensure that the Huawei Cloud account you added to the MgC Agent has the read-only permission for MRS and DLI. • If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed. – Under **Region**, select the region where the data to be verified is located. – Under **Project**, select the project where the data to be verified is managed. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. • If the selected credential is not the one you currently use to access MgC: – Under **Region ID**, enter the ID of the region where the data to be verified is located. For example, if the region is CN South-Guangzhou, enter **cn-south-1**. – Under **Project ID**, enter the project ID corresponding to the region. – Under **Cluster ID**, enter the ID of the cluster where the data to be verified is located. **NOTE** • To view the region ID and project ID, choose **My Credentials** > **API Credentials**. • For details about how to obtain the cluster ID, see **Obtaining an MRS Cluster ID**.

Step 13 After the connection test is successful, click **Confirm**. The cloud service connection is set up.

Step 14 Choose **Metadata Management** and click the **Tables** tab.

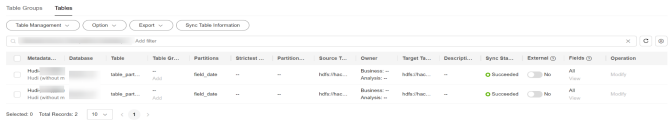


Step 15 On the displayed page, click **Import**.



Step 16 Click **Download** to download the import template to the local PC. Open the template, fill in table information, and save the template.

Step 17 In the **Import Tables** dialog box, click **Select File**, choose the filled template file, and click **Confirm**. After the import is complete, you can view the imported tables on **Tables** tab.



Step 18 In the **Features** area, click **Table Management**.

Step 19 Under **Table Groups**, click **Create**. Configure the **parameters for creating a table group** and click **Confirm**.

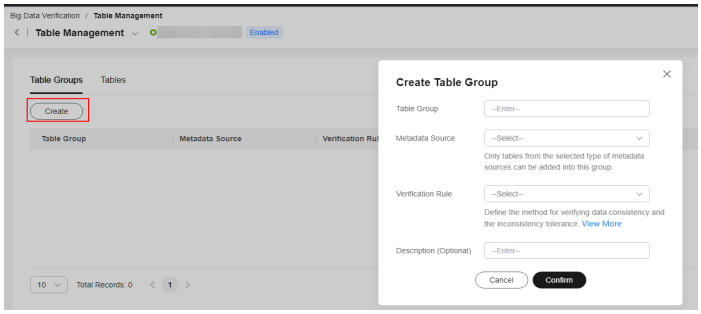
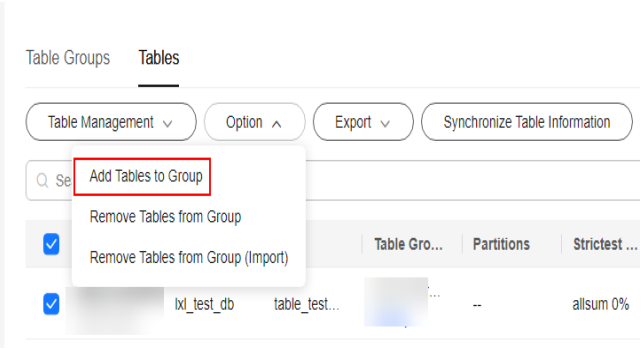


Table 7-45 Parameters for creating a table group

Parameter	Description
Table Group	User-defined
Metadata Connection	Select the connection created in **step 12**. **CAUTION** A table group can only contain tables coming from the same metadata source.

Parameter	Description
Verification Rule	Select a rule that defines the method for verifying data consistency and the inconsistency tolerance. MgC provides multiple verification rules for you to choose. For details about these rules, click **View More**.
Description (Optional)	Enter a description to identify the table group.

Step 20 On the **Table Management** page, click the **Tables** tab, select the data tables to be added to the same table group, and choose **Option** > **Add Tables to Group** above the list. In the displayed dialog box, select the desired table group and click **Confirm**.



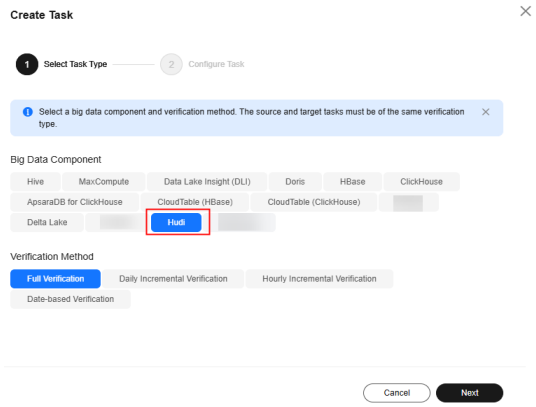
NOTICE

You can manually import information of incremental data tables to MgC. For details, see **Creating a Table Group and Adding Tables to the Group**.

Step 21 Create a connection to the source and target executors separately. For details, see **Creating an Executor Connection**. Select the source and target executor credentials added to the MgC Agent in **step 6**.

Step 22 Create a data verification task for the source and target Hudi clusters, respectively, and execute the tasks. For more information, see **Creating and Executing Verification Tasks**. During the task creation, select the table group created in **step 19**.

- On the **Select Task Type** page, choose **Hudi**.



- Select a verification method. For details about each verification method, see **Verification Methods**.

Step 23 Wait until the task executions enter a **Completed** status. You can view and export the task execution results on the **Verification Results** page. For details, see **Viewing and Exporting Verification Results**.

----End

8 Migrating Big Data Without Using the Internet

This section describes how to use NAT gateways and VPNs to migrate and synchronize big data when the MgC Agent has no Internet access. The following assumes that Alibaba Cloud is the source and the MgC Agent is installed on the Alibaba Cloud ECS.

Step 1: Configure a Huawei Cloud VPN

Step 1 Log in to the Huawei Cloud console and, in the service list, choose **Networking** > **Virtual Private Network**.

Step 2 Configure a VPN gateway

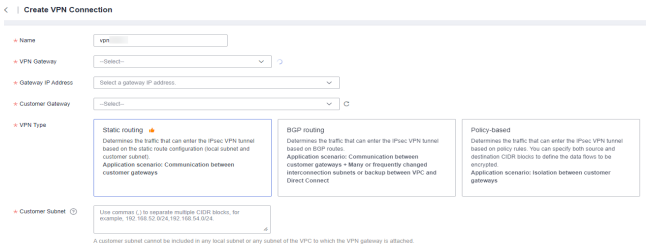
1. In the navigation pane, choose **Virtual Private Network** > **Enterprise – VPN Gateways**.
2. Click **Buy S2C VPN Gateway** and set parameters by following the on-screen instructions.
3. Configure all required parameters and click **Buy Now**.

Step 3 Configure customer gateways. You need to create an active and a standby customer gateway.

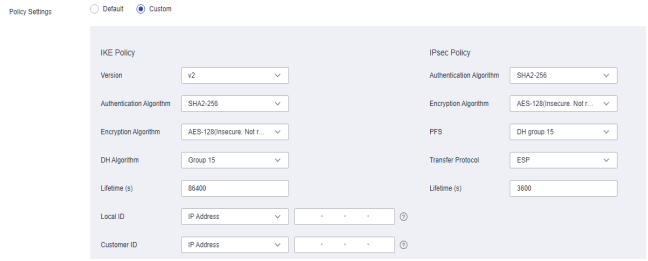
1. In the navigation pane on the left, choose **Virtual Private Network** > **Enterprise – Customer Gateways**.
2. Click **Create Customer Gateway** and set parameters by following the on-screen instructions. Select **IP Address** for **Identifier** and enter the public IP address of the Alibaba Cloud gateway.
3. Click **Create Now**.

Step 4 Create VPN connections. Create two VPN connections to connect to the Huawei Cloud VPN gateway and Alibaba Cloud customer gateway, respectively.

1. In the navigation pane on the left, choose **Virtual Private Network** > **Enterprise – VPN Connections**.
2. Click **Create VPN Connection**. On the displayed page, select the created VPN gateway and a customer gateway, and enter the subnet address of the customer gateway. Ensure that the subnet addresses do not overlap.



Select **Custom** for **Policy Settings** and ensure that the settings are the same as those on Alibaba Cloud.



3. Configure all required parameters and click **Buy Now**.

----End

Step 2: Configure an Alibaba Cloud VPN

Step 1 Sign in to the Alibaba Cloud console and choose **Products and Services** > **Networking and CDN** > **Hybrid Cloud Network** > **VPN Gateway**.

Step 2 Configure a VPN gateway

1. Click **Create VPN Gateway** and set parameters by following the on-screen instructions.
2. Configure all required parameters and click **Buy Now**.

Step 3 Configure the customer gateway.

1. In the navigation pane, choose **VPN** > **Customer Gateways**.
2. Click **Create Customer Gateway** and set parameters by following the on-screen instructions.
3. Click **OK**.

Step 4 Create a VPN connection.

1. In the navigation pane, choose **VPN** > **IPsec Connections**.
2. Click **Create IPsec Connection**, select the VPN gateway configured in step 2, and keep the policy settings the same as those on Huawei Cloud.
3. Click **OK**.

Step 5 Configure a route to the Huawei Cloud VPC subnet.

1. In the navigation pane, choose **VPN** > **VPN Gateways**.
2. Click the VPN gateway name. On the **Destination-based Route Table** tab, click **Add Route Entry** and set parameters based on the instructions.

----End

Step 3: Configure an Alibaba Cloud NAT Gateway

Create an Alibaba Cloud NAT gateway and configure SNAT and DNAT entries. For details, see **Creating and Managing an Internet Public NAT Gateway**.

Step 1 Sign in to the Alibaba Cloud console and choose **Products and Services** > **Networking and CDN** > **Hybrid Cloud Network** > **VPN Gateway**.

Step 2 Create an Internet NAT gateway.

1. On the **Internet NAT Gateway** page, click **Create Internet NAT Gateway** and configure parameters based on the instructions.
2. Configure all required parameters and click **Buy Now**.

Step 3 Configure an SNAT entry.

1. On the **Internet NAT Gateway** page, locate the Internet NAT gateway created in step 2 and click **Configure SNAT** in the **Actions** column.
2. On the **SNAT Management** tab, click **Create SNAT Entry** and set parameters based on the instructions.
3. Click **OK**.

Step 4 Configure a DNAT entry.

1. On the **Internet NAT Gateway** page, locate the Internet NAT gateway created in step 2 and click **Configure DNAT** in the **Actions** column.
2. On the **DNAT Management** tab, click **Create DNAT Entry** and set parameters based on the instructions.
3. Click **OK**.

----End

Step 4: Configure Security Groups

You need to configure security groups on Huawei Cloud and Alibaba Cloud.

Step 1 On the Huawei Cloud console, configure the involved security group to allow access from the private IP address of the server where the MgC Agent is installed.

1. Sign in to the Huawei Cloud console.
2. In the **Service List**, choose **Networking** > **Virtual Private Cloud**.
3. In the navigation pane, choose **Access Control** > **Security Groups**.
4. In the security group list, locate the security group where the target big data cluster is managed and click **Manage Rules** in the **Operation** column.
5. On the **Inbound Rules** tab, click **Add Rule**.
6. In the displayed dialog box, add a rule that allows TCP traffic to the open port of the MRS cluster. Enter the private IP address of the server where the MgC Agent is installed in the **Source** text box. For example, the default open port of a secured Hive cluster is 9083.
7. Click **OK**.

Step 2 On the Alibaba Cloud console, configure the involved security group to allow access from the private IP address of the server where the MgC Agent is installed.

1. Sign in to the Alibaba Cloud ECS console.
2. In the navigation pane, choose **Network & Security** > **Security Groups**.
3. Locate the security group that the server with the MgC Agent installed belongs to and click **Manage Rules** in the **Operation** column.
4. On the **Inbound** tab, click **Quick Add**. Set **Action** to **Allow**, **Authorization Object** to the public IP address of the server where the MgC Agent is installed, and **Port Range** to **All**.
5. Click **OK**.

----End

Step 5: Set Up a Migration Environment

Set up a migration environment by referring to **Preparations**. Purchase an ECS on Alibaba Cloud. Configure an SNAT rule for the NAT gateway to allow the ECS to access the Internet using its private IP address. Install the MgC Agent on the ECS, register an account, and **connect the MgC Agent to MgC**.

Step 6: Create a Big Data Migration Task

Review and understand the **precautions** about big data verification tasks. Perform the following steps to create a big data migration task:

Step 1 **Create a connection to MaxCompute**.

Step 2 **Create a connection to DLI**.

Step 3 **Create a metadata migration task** or **Create a data migration task** and execute it.

----End

